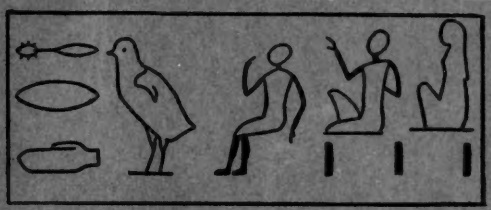
Vol. 8 SEPTEMBER, 1937 No. 3

CHILD DEVELOPMENT



Editorial Board

E. V. McCOLLUM
Professor of Biochemistry, School of Hygiene and Public Health, The Johns Hopkins University.

T. WINGATE TODD
Professor of Anatomy, Western Reserve University.

HAROLD E. JONES
Director, Institute of Child Welfare, and Professor of Psychology, University of California.

FREDERICK F. TISDALL
Director, Research Laboratories, Department of Pediatrics, University of Toronto, and Attending Physician, Hospital for Sick Children, Toronto.

MANDEL SHERMAN
The University of Chicago, Chicago, Illinois.

F. STUART CHAPIN
Professor and Chairman of Sociology and Director Graduate Social Work, University of Minnesota.

BUFORD JOHNSON, *Chairman*
Director of the Child Institute of the Psychological Laboratory, and Professor of Psychology, The Johns Hopkins University.

PUBLISHED QUARTERLY BY THE SOCIETY FOR RESEARCH IN CHILD DEVELOPMENT
NATIONAL RESEARCH COUNCIL
2101 CONSTITUTION AVENUE
WASHINGTON, D. C.
Made in United States of America

MATURATIONAL CHANGES IN RECTAL TEMPERATURES OF 61 INFANTS FROM 1 TO 36 MONTHS [1]

NANCY BAYLEY AND HERBERT R. STOLZ [2]

The children whose temperatures are here reported were normal, healthy infants who were brought to the Institute of Child Welfare repeatedly for a series of tests and observations. They came at regular intervals starting at the age of one month, every month during the first 15 months, at three-month intervals from 15 to 36 months, and at six-month intervals thereafter. At the visits during the first three years a number of physiological measures, including rectal temperature, were made. After three years of age the temperature readings were discontinued until six years, when mouth temperatures were taken.

The selection of the group and the procedures of observation have been described in some detail elsewhere (3). For the present purposes it may suffice to say that the children come, for the most part, from homes somewhat above average and have parents who are sufficiently interested to bring their children for the repeated visits. The testing procedures include mental and motor tests, physiological measures (blood pressure, breathing rate, pulse rate, and rectal temperature), anthropometric measurements, and photographs, usually presented in the order named. In the older children (after two years) the taking of temperature was usually postponed to the end of the examination period, as the children frequently objected to it. In order to maintain the best possible cooperation for all tests, the policy was followed of leaving to the last any procedures which might not be well received by the child. In the very young infants, on the other hand, when a child was brought in asleep, or nearly so, the physiological measures were sometimes taken first as they served to waken the child, yet were not usually disturbing to him before he was six months old. Otherwise they were taken after the mental and motor tests were completed and before the anthropometric measures were made.

Standard Fahrenheit rectal one-minute clinical thermometers were used and read after 1½ minutes. Mouth temperatures were taken at 72 and 84 months.

At each visit the mother was asked about the child's health and any illnesses which may have occurred since the last visit, and a pediatrician made notations of any conditions of the child's health which might give cause for concern. Starting at two years, a routine physical examination was made, and the pediatrician rated the child for nutrition, development, and physical handicaps to development. Each of these was rated separately, the first two on a seven-point scale, the third (where any handicaps were present) on a five-point scale. With a rating of "1" representing superior nutrition and development, and a "0" rating for no handicaps, a health score made by adding these three ratings together could range from 2 (excellent) to 19 (very poor), with a theoretical average health score of 10. Actually the range of scores was from 2 to 17 with a mean of approximately 8.4.

[1] Many of the computations used in this study were done on a W.P.A. project, No. 65-3-5406.

[2] Institute of Child Welfare, University of California.

RESULTS

Table I gives the mean rectal temperature and the standard deviation for all of the children measured, and for the boys and girls separately. The age curves of means, by sexes (Figure 1), show graphically that the temperatures tend to increase during the first seven months, remain constant from 7 to 24 months, after which they again drop.

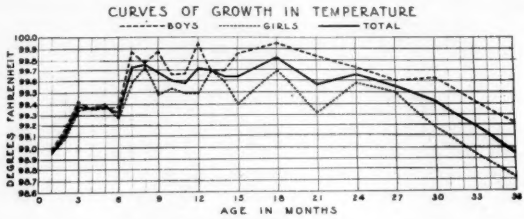


Figure 1. Curves of Mean Temperatures: Months 1 to 36

That this age trend of temperatures is a true characteristic of the group as a whole seems evident from means computed for fewer cases selected for comparisons within the group. When we divide the children by sexes, or when we select the 20 cases born in March to compare with 18 cases born in October and November, the means for all of these smaller samples show the characteristic increase of temperature during the early months, with a subsequent drop after two years of age.

Although sex differences are small and, as indicated by the SD's, overlapping is great, the boys show a consistent tendency to have slightly higher temperatures.

The dispersion of temperatures indicated by the SD's is also brought out in Figure 2, which gives percentage curves of the frequency of the occurrence of

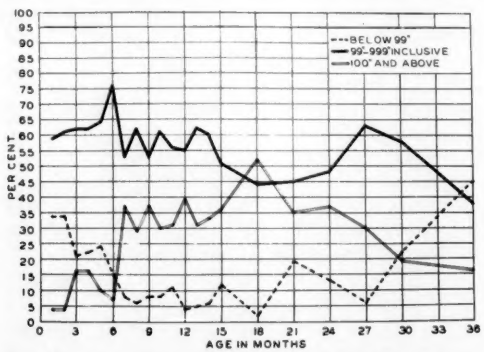


Figure 2. Percentages of Cases having High, Average, and Low Temperatures

Table I

Means and Standard Deviations of Rectal Temperatures

	Month	No. Cases	Mean	S.D.	P.E.
Boys	1	25	98.97	.96	
Girls		25	98.96	.16	
Total		50	98.96	.88	.08
Boys	2	31	99.16	.16	
Girls		25	99.09	.73	
Total		56	99.13	.36	.03
Boys	3	31	99.42	.70	
Girls		30	99.35	.24	
Total		61	99.38	.91	.08
Boys	4	29	99.36	.46	
Girls		28	99.37	.32	
Total		57	99.36	.80	.07
Boys	5	31	99.37	.96	
Girls		26	99.40	.32	
Total		57	99.38	.96	.09
Boys	6	30	99.38	.38	
Girls		27	99.27	.37	
Total		57	99.32	.96	.09
Boys	7	27	99.88	.53	
Girls		25	99.58	.60	
Total		52	99.73	.96	.09
Boys	8	26	99.78	1.00	
Girls		27	99.73	1.00	
Total		53	99.76	.68	.06
Boys	9	27	99.88	.82	
Girls		29	99.48	1.10	
Total		56	99.68	.53	.05
Boys	10	28	99.66	.63	
Girls		27	99.53	.97	
Total		55	99.60	.54	.05
Boys	11	26	99.67	.75	
Girls		24	99.49	.94	
Total		50	99.58	1.04	.10
Boys	12	25	99.95	.61	
Girls		24	99.49	.78	
Total		49	99.72	1.00	.10
Boys	13	23	99.70	.45	
Girls		20	99.70	.94	
Total		43	99.70	.72	.07
Boys	14	19	99.70	.65	
Girls		20	99.59	.56	
Total		39	99.64	.85	.09
Boys	15	23	99.86	.83	
Girls		21	99.40	.96	
Total		44	99.64	.95	.10
Boys	18	23	99.95	.68	
Girls		20	99.70	.61	
Total		43	99.83	.89	.09
Boys	21	22	99.84		
Girls		24	99.32		
Total		46	99.57	.78	.08
Boys	24	19	99.73		
Girls		15	99.58		
Total		34	99.66	1.03	.12
Boys	27	15	99.60		
Girls		15	99.50		
Total		30	99.55	.50	.06

Table I (continued)

	Month	No. Cases	Mean	S.D.	P.E.
Boys	30	22	99.62		
Girls		18	99.17		
Total		40	99.42	.69	.07
Boys	36	15	99.22		
Girls		19	98.72		
Total		34	98.95	.66	.08

deviating temperatures as compared with the average - those below 99.0 degrees, those from 99.0 to 99.9 degrees, and those 100 degrees and above. The curve of the central tendency has obviously been influenced by the age differences in extremes of temperature. Readings below 99.0 degrees are rare between 6 and 21 months, while readings of 100 degrees or slightly higher are comparatively frequent during these same ages.

The question arises whether this change in temperatures with age is a true physiological process of maturation or is brought about by the conditions under which the temperatures were taken. As a rule children's temperatures reported in the literature have not been given in such a way that any age changes might be detected. Benedict and Talbot (6) give norms for rectal temperatures for comparable numbers of cases at most months up to one year and at less frequent intervals through twelve years. They report mean values ranging, for boys from 97.1° F at five years (only two cases to 99.5° at 1½ years (6 cases). The mean values for girls range from 98.3° (30 cases) on the first day of life to 99.5° at 11 months (16 cases) and at 1½ years (8 cases). In general they found high average temperatures for boys from ten months to two years and for girls from eleven months to two years. These curves show similar trends to ours except that the temperatures do not rise as high (they were taken during basal metabolism experiments) and the rise occurs three to four months later than in our group. Their results may differ from ours in part because they omitted all temperatures over 100° F.[1]

Our readings were taken between 10 a.m. and 5 p.m., as indicated in Table II, which is based on the total number of readings at eight representative ages (months 1, 3, 9, 12, 18, 24, 30, and 36). Since it was our policy to make the appointments for a child always at the same time of day, individual children's fluctuations in temperature are seldom due to variations in the time of day the readings were taken, and there is no piling up of early or late readings at some ages as compared with others. Table II shows that there is a tendency for higher readings in the afternoon than in the morning. These findings corroborate earlier studies (2, 12, 15, 19), but they do not account for the age changes in the means for this group.

Another extraneous factor which may have affected the temperatures is the emotional disturbance caused by the testing situation. It was necessary to handle the children considerably during the procedures of testing and measuring, and it was thus inevitable that some of them should be upset (4). Also, fearfulness of the strange situation was evident after the first half year. If strong emotions tend to raise the temperature (17), the higher means from months 7 to 24 might be

[1] Jacobsen, Jacobsen & Yoshioka (13) report daily temperatures for an infant chimpanzee, Alpha, from birth to one year. Her temperatures show a tendency to increase from birth to one week and from three to six months of age, with a drop starting during the seventh month and lasting through the ninth.

Table II

Temperatures for Eight Ages Combined, Classed according to Time of Day

Hour	10:00	11:00	12:00	1:00	2:00	3:00	4:00
No. of Cases	71	60	12	17	64	80	65
Mean Boys	99.49	99.42	....*	....*	99.83	99.71	99.98
" Girls	98.96	99.13	....*	....*	99.65	99.57	99.76
" Total	99.22	99.27	....*	....*	99.74	99.64	99.87
S.D.Total	4.44	3.18			2.78	3.02	2.70

*Means were not computed at 12:00 and 1:00 because of the small number of cases measured at these hours.

a result of more emotional disturbance at these ages. Since crying is one objective measure of emotional disturbance in an infant, the amount of time he cried during the test procedures was recorded, and its percentage of the total examination period computed. The median percentages for the group show that the amount of crying diminishes from 1 to 4 months, then increases to 12 months, after which age it diminished rather rapidly, occurring in only a few cases after 24 months of age. The only period where the trends of the mean temperature and crying curves parallel each other is between 6 and 12 months. Coefficients of correlation were computed between temperature and percent of crying at months 1, 3, 6, 9, 12, 18, 24, and 36. The r's are all zero (a representative r being -.09) except at month three where it is .39 $\pm$ .08. A comparison of the curves of crying and temperature for individual children showed that the two curves rarely parallel each other. Higher temperatures during examinations when a child cried more than usual do not appear more often than would be expected by chance. In a few cases both high temperature and crying seemed obviously to be related to some disturbance in the child's health at the time of the test. It appears, then, that the age trends obtained for temperatures in this study are not due to the effects of time of day or of emotional disturbance of the child.

To determine whether the variations might be seasonal, depending on the birth month, means were computed separately for the twenty cases born in March, and for eighteen cases born in the fall, mostly in October and November. In both of these samples the means are closely similar to each other and to the means for the total group. The season of the year at which the child was born appears to have no influence on the trend of his temperatures.

STABILITY OF TEMPERATURE IN INDIVIDUAL CHILDREN

There are so many factors which do affect the temperatures of healthy children --diurnal rhythm, sleep, rest and activity, ingestion of food--which could not be held constant in this study, that we should expect considerable variation from month to month in the temperature readings of the same child even though there were no tendency for temperatures to change with age. The fact that coefficients of correlation computed between the temperatures read at two successive months are all around zero may be due either to these diurnal variations or to individual differences in rates of growth changes. In a few cases readings were repeated under more controlled conditions. Figure 3 gives graphically a healthy six-year-old girl's temperatures which were read several times a day for 21 consecutive days. Her morning readings are seen to be, in general, lower than in the afternoon, but for any one hour of the day there is considerable variation. The eight

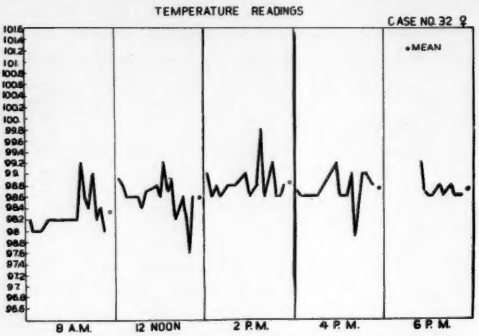


Figure 3. Daily Temperatures of Case 32 at Six Years

a.m. temperatures range from 98.0 to 99.2 degrees; at ten o'clock they range from 97.6 to 99.2 degrees; at two o'clock, from 98.6 to 99.8 degrees; and at 4 o'clock, from 97.9 to 99.2 degrees. It is small wonder, then, that the children's temperatures taken at intervals of a month or longer do not remain constant.

It has seemed probable that variability in temperature is greater in young infants and that as they mature their physiological functions (including temperature regulation) will become more stable. Although the standard deviations from the means show little evidence of such a tendency throughout the first two years, the 27, 30, and 36 month readings are less variable than at most of the earlier ages. And, in addition, the mouth temperatures taken at 72 and 84 months (Table III) have about the same SD's as these later rectal temperatures. Individual children's temperatures may be growing less variable even though the standard deviations for the group remain fairly constant. Children may differ in the ages at which their temperatures are becoming higher or lower, as well as in the amount of fluctuation in temperature which is normal for them.

Table III

Means and Standard Deviations of Mouth Temperatures at Months 72 and 84

	Month 72			Month 84		
	No. Cases	Mean	S.D.	No. Cases	Mean	S.D.
Boys	11	99.2		19	98.8	
Girls	14	99.2		20	98.8	
Total	25	99.2	.56	39	98.8	.59

Since the average curve for the group shows increasing temperatures from 1 through 7 months, and decreasing temperatures after 24 months, we may expect that most children would exhibit the same tendency as they mature from birth through three years. An inspection of the individual curves shows this to be true, in general, with individual variations in the age at which the higher temperatures occur. However, the fluctuations in temperature resultant from the immediate conditions of environment, diurnal rhythm, etc. make the individual curves very irregular. It may be permissible to assume that these irregularities are variations

on a central tendency which is more fundamental, and related to the child's age and inherent physical make-up. Such an assumption can be made only with strong reservations because of the infrequent intervals at which the temperatures were read. However, the following discussion is based on this assumption.

If we wish to study variability in individuals, a child's deviation from his own central tendency should take his age trends into account. In order to do this we plotted curves of the individual children's temperatures with age and drew (by eye) a smoothed line through each curve. Assuming this smoothed line to be indicative of the central tendency of temperature for that child, the deviation of the actual reading from this line for each age was read from the chart and recorded as his "deviation score." The means of these deviation scores were computed for months 2 to 15 inclusive and for months 18 to 36 inclusive. As is shown in Table IV the total group averages of these deviations are no larger for the first 15 months than for the 18 to 36 month period. The average deviations of these deviations, however, are smaller for the later age group, indicating a slightly greater stability with increasing age.

Table IV

Means and Average Deviations of Deviation Scores, Younger Ages Compared with Older Ages*

	Months 2 to 15		Months 18 to 36	
	Mean	A. D.	Mean	A. D.
Boys	.32	.17	.31	.12
Girls	.37	.19	.37	.14
Total	.34	.18	.34	.13

*The means for the separate ages are not given, but instead, the mean of the mean deviation scores for months 2 to 15 inclusive, and another mean for months 18 to 36, inclusive.

Some children appear to be far more labile than others, and there are wide differences in the extent to which a child's temperature fluctuates. This is shown in Figures 4 and 5 where the highest and the lowest readings taken for each child are indicated. Figures 6 and 7 illustrate this in another way with curves of individual children, two with small variation in temperatures and two with wide variation.

The age curve of the mean temperatures, up to a certain point, follows the curve of body build (Weight divided by Length, squared) for these same children (5). The mean W/L^2 index rises from 1.41 at month one to 1.83 at month eleven, then drops until at month thirty-six it is 1.63. At seven months, when the temperatures have reached an approximate plateau around 99.7^0, the W/L^2 index has gone up to 1.76; by twenty-four months the index has returned to 1.71 and after this the temperatures drop as the children become more slender. At the ages when children are most chubby their temperatures tend to be higher. During the period of increasing chubbiness (W/L^2) some children become much more chubby than others. Possibly a child's temperature tends to be high or low according to his body build. That is, plump, relatively heavy children may as a rule have higher temperatures than thin, relatively light children. When we try to compare these two variables, we are confronted, in both instances, with data which fluctuate widely. Temperature varies with time of day and is more affected by immediate conditions; while the W/L^2 index varies over somewhat longer intervals with variations in rates of growth.

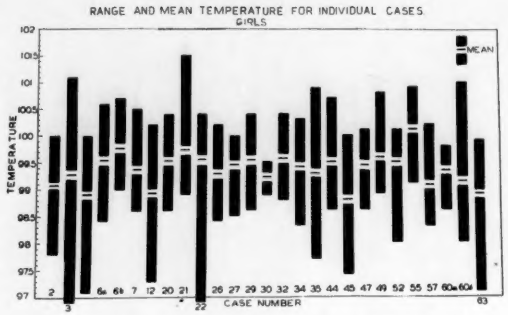


Figure 4. Range and Mean Temperature for Individual Cases: Girls

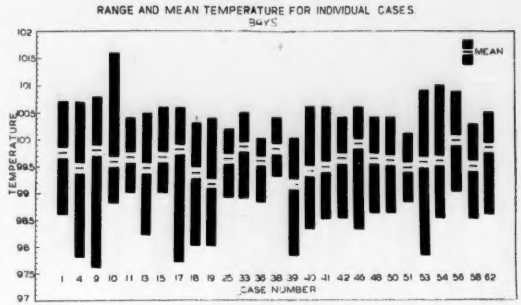


Figure 5. Range and Mean Temperature for Individual Cases: Boys

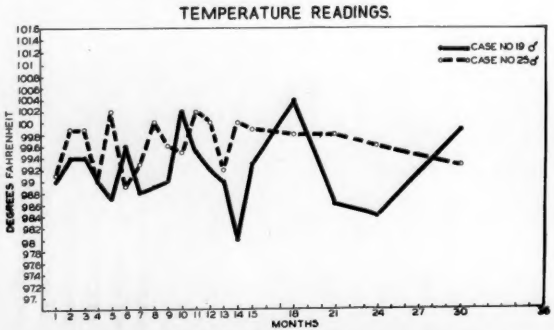


Figure 6. Individual Curves of Temperatures

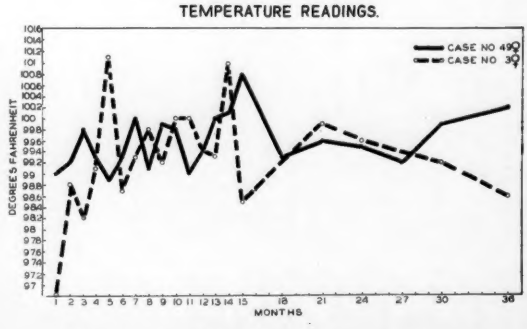


Figure 7. Individual Curves of Temperatures

From a comparison of individual curves of the W/L^2 index and temperature readings, it was evident that no correlation would be found for any one age between a child's temperature (single reading) and his W/L^2 index. However, when a mean temperature (composed of all temperatures taken at all ages when the child was well) was computed for each child, this average measure correlated with the W/L^2 index at 12 months $+.35 \pm .087$, and at 36 months $+.42 \pm .080$. Although these r's are not high, they are greater than would be expected by chance. They appear to indicate that body build is one factor which is related to the child's temperature.

THE RELATION OF TEMPERATURES TO HEALTH AND ILLNESSES

Donald (10), in a study of twenty institution children who were carefully selected as in excellent physical condition, found a tendency toward constant high temperatures in 30% of the cases, and a few children whose temperatures were characteristically "subnormal." Abt (1) reports similar findings from his practice.

As previously stated the temperature readings used in this study were made on presumably well children. However, these children present some differences in their health records; and the extreme high or low temperatures found at times when the children were apparently well may be related to poor health or a tendency toward frequent illnesses.

In order to study the relationship to illnesses, all illnesses were rated for their severity, on the basis of the mothers' descriptions and the nature of the disease, on a five-point rating scale. Such things as a slight cold were given a rating of "1"; severe illnesses, a rating of "5". In this way we were able to compute an illness score (incidence weighted for severity) for each child for any age-interval which it was desired to study. Figure 8 gives the incidence, in three-month intervals, of the infectious diseases occurring in these children during the first three years. Colds and infections other than epidemic diseases were found to occur at all ages. While epidemic diseases did not put in an appearance during the first six months, they were present in small numbers throughout the rest of the age-range. When all illnesses were rated for severity, the total weighted incidence for each three-month interval was found to be as represented in Figure 9. Comparing this figure with the curve of mean temperatures, we find some correspondence between the ages of more illness and high temperatures, though there are exceptions at 18 and 36 months. When each child's average temperature for all readings through 36 months is compared with his weighted illness score for the total 36-month period, the correlation is $.28 \pm .087$. This may indicate a slight tendency for high temperatures to be found in the children with more frequent illnesses, as well as at the ages when illnesses were more

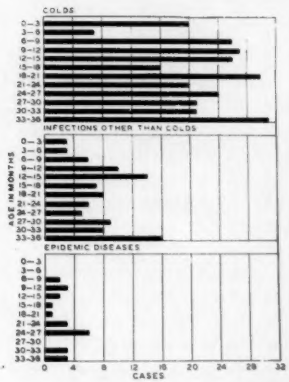


Figure 8. Incidence of infectious Diseases.

frequent. This relation, however, is not statistically significant.

When we correlate the average temperatures with the health scores based on the physician's ratings, the r with the score at 21 months is +.21 and at 36 months, -.22. In these children present health bears no relation to an average of past temperatures.

We may study this health-temperature relationship further by considering the children whose temperatures were repeatedly 100 degrees or above and those with temperatures repeatedly below 99.0 degrees. To make this comparison we carefully excluded all readings made at a time when a child showed any other evidence of infection, digestive upset or other physical ailment which might possibly affect the temperature. With the exception of two girls all of the children had at least one reading of 100 degrees or above and many had three, four, or five such readings. There were five boys and one girl whose temperatures, under these conditions, exceeded 99.9 degrees for 50% or more of the measures. They are given in Table V.

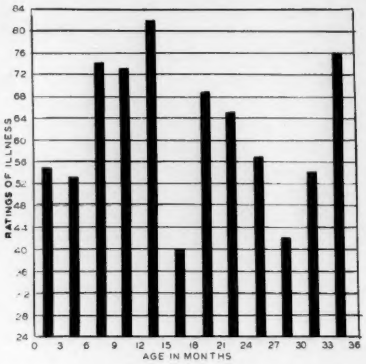


Figure 9. Weighted Incidence of Illnesses by Three-Month Intervals.

Table V

Cases with Frequent High Temperatures

Case	Temperatures 100° or above	Per Cent	Usual Time of Day
9	7 of 12	58	9 to 10 a.m.
17	7 of 13	54	2 to 3 p.m.
33	12 of 16	75	3 to 4 p.m.
46	12 of 20	60	3 to 4 p.m.
56	10 of 17	59	10 to 11 a.m.
55	14 of 18	78	3 to 4 p.m.

Four of these children were usually tested in the afternoon, when temperatures tend to be higher, but the proportion of morning to afternoon readings for these six children is the same as for the entire sample. Cases 9, 17, 46, and at some ages 56 have a chubby body build. Three of the children (cases 9, 55, and 56) had high illness scores, while a fourth (case 33) had a rather high illness record. Of these three factors which have been shown to be more or less related to high temperatures, at least one, and usually two, are found for each of these children.

Cases 9 and 56 attended the nursery school, and we have for them a series of daily temperatures taken at the same time of day when they were about three years of age.[1] Case 9, for ten daily readings at 9:00 a.m., had a mean temperature of 99.6° with a range from 98.9° to 100.1°. For another series of ten days his nine o'clock readings ranged from 99.0° to 99.8° with a mean of 99.5°; however, one

[1]These figures were kindly supplied by Mrs. Gladys N. Ludwig of the Institute Nursery School.

hour later on these same ten days his range was from 100° to 101.1° with a mean of 100.6°. His temperatures appear to be very easily raised by exercise and active play.

Case 56, at 9:00 a.m., had ten temperatures ranging from 99.7° to 100.5° with a mean of 99.9°--definitely above the mean for the entire nursery school group of 99.4° at 9:00 a.m.

Although most of the children had some readings below 99.0°, only one child had such low ratings as often as 50% of the time, and there seems to be no significant relation between illnesses and frequent low temperatures.

This approach shows again that the children in this group who were prone to high temperatures when they were well were also, as a rule, more subject to illnesses. But the exceptions indicate that considerable variation in temperatures may normally be expected in healthy children. (Several children with high illness scores do not fall into the high-temperature group.)

SUMMARY

In summary, an analysis of the rectal temperatures taken on this group of 60 children, as they grew from birth to three years, has brought out several points.

1. Temperatures in infants tend to increase during the first seven months and to drop again between 24 and 36 months of age. The mean rectal temperatures for the total group vary for the different ages from 98.96° at month one to 99.76° at month eight and 99.83° at month eighteen, and drop back to 98.95° at month thirty-six.

2. Mouth temperatures for 25 children at six years averaged 99.2°, and for 39 children at seven years averaged 98.8°.

3. The boys' temperatures are, on the average, slightly higher than the girls' although the overlapping in the sex distribution is very great.

4. The temperatures taken varied widely when we take into account what is usually considered "normal". This is true for the group at any one age, and for any one child at various ages. The variability appears to be somewhat less after the first two years.

5. Aside from changes due to maturation and diurnal rhythms, these children's temperatures appear to be related to a number of factors. Among these factors are body build and possibly susceptibility to illnesses.

6. There is some evidence that healthy children may have consistent tendencies toward high temperatures or toward low temperatures which are normal for them.

REFERENCES

(1) Abt, J. A.: Temperature variations in infancy and early childhood. Illinois Med. J., 1919, 36, 5-10.

(2) Bardswell, N. D. and Chapman, J. E.: Some observations upon the deep temperature of the human body at rest and after muscular exertion. Brit. Med. J., May 13, 1911, 1106-10.

(3) Bayley, N.: Mental growth during the first three years: A developmental study of sixty-one children by repeated tests. Genet. Psychol. Monog., 1933, 14, 1-92.

(4) Bayley, N.: A study of the crying of infants during mental and physical tests. J. Genet. Psychol., 1932, 40, 306-329.

(5) Bayley, N. and Davis, F. C.: Growth changes in bodily size and proportions during the first three years: A developmental study of sixty-one children by repeated measurements. Biometrika, 1935, 27, 26-87.

(6) Benedict, F. G. and Talbot, F. B.: Metabolism and growth from birth to puberty. Carnegie Inst. of Wash. Publ., #302, 1921, pp. 213.

(7) Benedict, F. G. and Talbot, F. B.: The physiology of the newborn infant. Carnegie Inst. of Wash. Publ., #233.

(8) Burton, A. C. and Murlin, J. P.: Human calorimetry. III. Temperature distribution, blood flow and heat storage in the body in basal condition and after ingestion of food. J. Nutrition, 1935, 9, 281-299.

(9) DeBuys, L. R., Bacal, H. L. and Stewart, A. M.: Elevations of temperature in the newly born. Am. J. Dis. of Child., 1933, 45, pp. 520.

(10) Donald, W. M.: Some observations upon the temperature of apparently healthy children; an experimental study. Arch. Pediat., 1901, 18, 189-192.

(11) Fraenkel, Dora: Über die normale Korpertemperature der Kinder and ihr Verhalten bei Bewegung und Ruhe. Deutsche Med. Wchnschr., 1913, 39, 267-268.

(12) Funk, F. H. and Gordon, B.: Body temperature in pulmonary tuberculosis. Am. Rev. Tuberc., 1927, 16, 692-699.

(13) Jacobsen, C. F., Jacobsen, M. M. and Yoshioka, J. G.: Development of an infant chimpanzee during her first year. Comp. Psychol. Monog., 1932, 9, 1-94.

(14) Kerby, C. J.: Observations on temperature in children. Pediatrics, 1907, 19, 521-529.

(15) Ludwig, G. N. and Stolz, H. R.: Temperature variations in normal preschool children. Child Development, 1931, 2, 225-228.

(16) McConnell, W. J. and Houghton, F. C.: Body temperatures and their measurements. J. Am. Soc. Heat & Ventil. Engin., 1921-22, 28, 523-532.

(17) Neff, F. C.: Temperature variability in certain apparently normal children. Southern Med. J., 1922, 15, 268-272.

(18) Rautmann, H.: Untersuchungen über die Variabilität der Körpertemperatur bei Gesunden. Ztschr. f.d. Ges. Anat. (Abt. 2), 1928, 13, 588-592.

(19) Schaefer, E. A. S.: Textbook of Physiology (Pembrey, M. S.), 1898, Vol. 1, page 785.

TWIN RESEMBLANCES IN MECHANICAL ABILITY, WITH REFERENCE TO THE EFFECTS OF PRACTICE ON PERFORMANCE [1]

DAVID BRODY [2]

INTRODUCTION

Although extensive studies on twin resemblances have served as an important approach to the problem of mental inheritance, no twin data relating to the inheritance of mechanical ability have been reported in the literature. The present study aims at an approach to this problem through an analysis of the degree of twin resemblances in performance on the Minnesota Spatial Relations test which is a form board designed to measure mechanical ability.

Of importance in an evaluation of the twin technique is the assumption that the environment of fraternal twins of like sex is as uniform as that of identical twins. Such an assumption requires close examination, for unless we can assume a constancy of environment for both fraternal and identical twins, any inferences as to genetic contributions must be modified accordingly. For one thing, the very factor of genetic dissimilarity will tend to operate so as to produce a greater variability of environment for fraternal twins than for identical pairs. Stocks (11) illustrates this point very well when he states that, ".....many dizygotic twins are very different in general body build, healthiness, taste, and temperament, so that they naturally tend to subject themselves or be subjected to differences in nurture to a greater degree than monozygotic twins who have usually the same needs, tastes, and inclinations, and are rarely seen apart during childhood."[3] Likewise Jones and Wilson (5) in a study of reputation differences between like sex twins find that individual pairs of monovular twins are generally considered to be more alike than diovular pairs and conclude ".....that this effect is to contribute an environmental differentiating factor which is greater for fraternal than for identical twins."[4] It becomes important, then, to recognize the role of inheritance in affecting the degree of similarity of environment between twin pairs.

One may, however, seriously question whether the greater similarity of environment of identical twins is sufficient in itself to account for the marked differences in the degree of resemblance between the two types of twins. The point to be emphasized, rather, is the need of caution in drawing inferences from comparative studies of identical and fraternal twins.

Diagnosis of Twinning

The writer in his work employed a scheme which combined the criteria taken from the separate scales constructed by Dahlberg (1), Siemens (10), and Newman (8). These criteria are:

[1] The writer wishes to express his appreciation to Dr. Florence L. Goodenough and Dr. John E. Anderson for helpful suggestions and criticisms.

[2] From Institute of Child Welfare, University of Minnesota.

[3] Stocks, P. Annals of Eugenics. April, 1930, page 104.

[4] Jones, H. E. and Wilson, P. T. J. Exper. Educ., 1932-33, 1, p. 91.

1. The appearance of the twins must give an impression of very great resemblance or identity.
2. That during childhood, neighbors, school fellows, etc. have had difficulties in distinguishing them and have sometimes confused them.
3. That the configuration of the ears does not show great dissimilarity.
4. That the finger prints show a high degree of similarity.
5. That the anthropological measurements do not show any marked difference.

Measurement of Mechanical Ability

The test used was the Minnesota Spatial Relations Test which was selected from the battery devised in the Minnesota Mechanical Ability Studies (9). An examination of the data relating to the separate tests in this battery indicates that of the tests listed the Minnesota Assembly and the Minnesota Spatial Relations tests most adequately meet the criteria of validity, reliability, and uniqueness. The significance of uniqueness is best indicated by the following quotation from the Minnesota study:

> ".....From the point of view of the theory of unique traits the task of measuring mechanical ability must meet two theoretical demands. The first is that the tests be constructed to correspond with actual achievements in work generally regarded as requiring mechanical ability, identifying them as tests of mechanical ability only on the basis of correlations with outside criteria. The second demand is that the measures be of mechanical ability and not of some other ability which may be closely related to it and which for that reason might possibly be confused with it."[1]

A study of sex differences points to the Spatial Relations test as the more satisfactory of the two for a research study of this kind. Strikingly greater sex differences on the Assembly test strongly suggest the effect of environmental factors upon performance in this test. The Minnesota study (9) indicates that the superiority of the males must be attributed to social pressures which allow them greater opportunity to engage in the specific type of mechanical work measured by the assembly test. The spatial relations test which is just as valid a measure of mechanical ability (as determined by the shop-work criterion) fails to indicate a distinct male superiority, and hence, suggests that opportunity for specific practice is considerably less for this type of performance.

The Minnesota Spatial Relations test consists of four form boards containing 58 cut-outs each. The test is so constructed that Boards A and B contain identical cut-outs differing from one another only to the extent that the recesses of each board are arranged in a different pattern. Boards C and D are constructed in a similar manner, but the cut-outs used are different than those for Boards A and B. The entire test, then, can be subdivided into two comparable tests: (1) Boards A and B; (2) Boards C and D.

The reliability of the four boards when administered to 100 Junior High School boys as reported in the Minnesota Study was .84. When checked against a shop-work

[1] Paterson, D. G. et al. Minnesota Mechanical Ability Studies, page 22.

criterion, the test yielded a validity coefficient of .53.

PROCEDURE

In a survey of the Minneapolis public schools, sixty-two pairs of twin boys of grade school age were secured for study. Owing to the difficulty of securing twin subjects, it was found necessary to accept all available twins in the grade school population. The distribution of these pairs according to chronological age is indicated in Table I.

Table I

Distribution of Twin Pairs According to Chronological Age

C.A.*	Number of Pairs		
	Identical	Fraternal	Total
8	5	7	12
9	5	6	11
10	1	5	6
11	6	6	12
12	8	6	14
13	3	3	6
14	1	0	1
TOTAL	29	33	62
AV. C. A.	10.78	10.13	10.44
S. D.	1.78	1.62	1.73

*Chronological age is taken from the nearest half year. Thus, all children varying in age from 7 years, 6 mos. to 8 years, 6 months, are classified under the 8 year category.

The testing program was carried out over a period of eight months, from January to September, 1935. During the winter and spring months the subjects were tested at the schools. In the summer, they were tested at their homes if conditions permitted; otherwise, they were brought to the Child Welfare Institute of the University of Minnesota.

The general procedure as outlined in the report of the Minnesota mechanical ability studies (9) was followed in detail in the administration of the test.

In the cases discussed here, a high degree of motivation was maintained throughout the testing period by encouraging each subject to improve his previous score. In addition, intra-pair competition was stressed so that each twin was constantly motivated by the desire to excel his brother. Rapport was established with very little difficulty, since the boys regarded the test as a contest or a game.

In order to determine the effect of practice on twin resemblances each subject was given six successive trials on Boards A and B in alternation. These trials constituted the practice series. Immediately following the practice series each subject was given the final test on Boards C and D. Each subject proceeded with the subsequent trial as soon as he indicated that he was sufficiently rested. In general, the intervals between tests varied from one to five minutes.

Wherever possible the testing of one member of a pair was immediately followed by the testing of the other member. In all cases both members of a pair were tested during the same day.

Scores are expressed in terms of the total number of seconds taken to place the 58 blocks in each board.

In order to determine whether any relationship exists between the form board performance and intelligence in children younger than those on whom the Minnesota test was first standardized, the Kuhlman-Anderson group tests were administered

to all twin pairs. The intelligence tests were administered within one to two days after the subjects had been given the Spatial Relations test.

RESULTS

Inter-correlation of Test Scores

The reliabilities of Boards A and B and of Boards C and D as determined by correlations between tests are presented together with the mean score and the standard deviation for each board in Table II. In computing the reliability of Boards A and B only the first trial on each was considered. The greater variability of scores on Boards A and B makes the comparatively larger reliability coefficient of Boards C and D still more significant.

Table II

Reliability of the Test

	Mean	S.D.	Reliability	Reliability (Sp.-Br.)
Board A	494.7	163.1		
Board B	391.3	129.2	.744	.853
Board C	397.9	137.1		
Board D	370.8	121.5	.764	.885

The increased reliability may be due to practice effects. Darley (2), who administered the test to adult subjects, reports greater reliability for Boards B, C, and D when Board A is used as a practice trial.

In a study of practice effects it becomes important to determine whether the same function is being tested from one practice trial to another. In the test used in this study it is possible that the later trials measured a memory factor unrelated to mechanical ability.

Evidence that the practice trials measured mechanical ability may be found in an inspection of the correlations between the practice trials and the combined scores in the final series. These data, presented in Table III, indicate a consistently high relationship between practice trial performance and the combined scores on C and D.

Table III

Correlations between the Scores on the Separate Practice Trials and the Combined Score on Boards C and D

Practice Trial	Correlation*
1A	.789
2B	.843
3A	.817
4B	.835
5A	.858
6B	.871

*The P.E. r for r = o is $\pm$ .06

Since the arrangement and the forms of the cut-outs in Boards C and D are different from those in Boards A and B, it cannot be assumed that the high relationship is due to practice in memory-motor functions unrelated to mechanical ability. Rather, the evidence suggests that a similar type of performance is being measured from one practice trial to the other.

Effects of Practice

The data presented in Table IV show the relationship of age and practice to performance. Because of the small number of cases at each age level the scores

Table IV

Mean Scores Made by Subjects of Different Age Levels

	8 yr. (N=24)	9 yr. (N=22)	10 yr. (N=12)	11 yr. (N=24)	12 yr. (N=28)	13 yr.* (N=14)
1-A	593	569	552	443	416	314
2-B	494	467	429	346	317	287
3-A	377	391	320	296	256	246
4-B	361	359	352	263	260	225
5-A	343	334	297	262	241	205
6-B	328	338	285	252	245	210
C	485	500	437	357	313	292
D	445	436	402	346	311	278

* In this instance the two members of the 14 yr. old pair have been classified under the 13 yr. category.

of the various age groups were combined into two categories, namely, an 8 to 10 yr. group and an 11 to 14 yr. group. The curves for these two groups and that for the total number of subjects are plotted in Figure 1. In all three cases the curves tend to approximate the general form of a negatively accelerated curve.

A study of the curves relating to performance on Board C indicates a transfer effect carried over from performance on the practice series. Thus, for all subjects there was an average reduction of 86 seconds in time scores between the first practice trial and the performance on Form C. This improvement of score was also characteristic for each of the age groups, and is indicated by the magnitude of the critical ratios presented in Table V.

Table V

Differences in Mean Scores on Trials 1A and Form C

Age Level	Av.-1A	Av.-Form C	Diff.	σ diff.	D/σ diff.
8	593	485	108	6.6	16.4
9	569	500	69	8.1	8.5
10	552	437	115	9.6	11.9
11	443	357	86	6.8	12.6
12	416	313	103	6.3	16.3
13	314	292	22	5.6	3.9
8-10	575	481	94	6.1	15.4
11-13	404	325	79	5.0	15.8
ALL	484	398	86	5.2	16.5

There was a general tendency for variability of scores to decrease with practice. This tendency is illustrated in Table VI which indicates that variability is at a minimum on the fifth trial and increases slightly on the sixth.

Table VI

Mean Scores and Standard Deviations

	* 8 - 10 (N=58) Mean	S.D.	11 - 14 (N=66) Mean	S.D.	All (N=124) Mean	S.D.
1A	575	145	404	133	484	163
2B	470	128	321	80	391	129
3A	370	88	268	47	316	86
4B	353	76	253	51	300	81
5A	330	72	241	44	283	74
6B	323	80	240	49	279	77
C	481	141	325	80	398	137
D	432	121	316	92	370	121

* Age taken from the nearest half year.

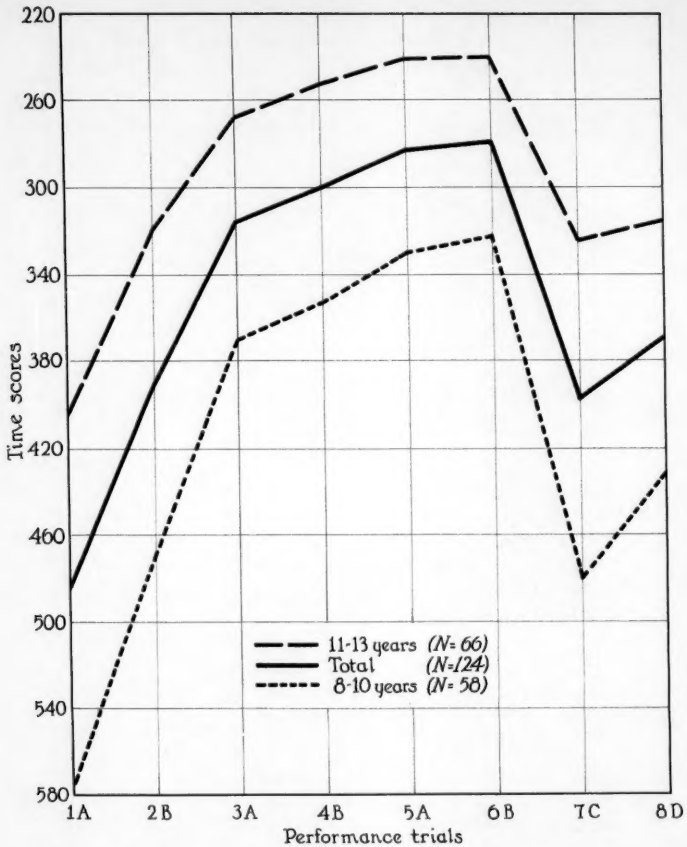


Figure 1.- Mean Performance Scores by Age Groups.

However, any comparison of the mean scores and standard deviations of the separate age groups and of the practice trials must be regarded very carefully, since the scores as indicated by the time units cannot be assumed to be equal at different points on the practice curve. Thus, a reduction of time from 240 to 230 seconds probably represents a greater improvement than a reduction of time from 530 to 520 seconds. That is, the score units which are "physically" comparable cannot be regarded as "psychologically" comparable.

Although the members of the younger group are more variable than the older subjects, the coefficients of variability in Table VII suggest that this greater variability may be a product of the unequal units of the learning curve.

Analysis of Intra-pair Differences

A comparison of the means of the absolute differences in scores between the

Table VII

Coefficients of Variability

	8 - 10	11 - 14
1-A	25.2	32.9
2-B	27.2	24.9
3-A	23.8	17.5
4-B	21.5	20.2
5-A	21.8	18.2
6-B	24.8	20.4
C	29.3	24.6
D	28.0	29.1

twin pairs as presented in Table VIII shows a markedly closer resemblance for the identical twins. This greater similarity of identical pairs becomes all the more evident when an analysis is made of the means of the intra-pair differences on the combined scores of Boards C and D. In this case a D/σ diff. value of 12.6 was obtained. The inconsistent trend in the size of the critical ratio with practice suggests that practice as such has no effect on the degree of differences in resemblance between fraternal and identical twins.

Table VIII

Means and Standard Deviations of Absolute Differences between Pairs Together with Critical Ratios

	Fraternal ($N=33$)				Identical ($N=29$)				Diff. Between Means	$\sigma_{diff.}$	$D/\sigma_{diff.}$
	Mean	σ_{mean}	S.D.	$\sigma_{s.d.}$	Mean	σ_{mean}	S.D.	$\sigma_{s.d.}$			
1-A	108.3	13.1	75.2	9.3	86.4	15.9	85.4	2.9	21.9	5.38	4.1
2-B	97.7	15.3	87.7	10.8	60.5	9.8	53.0	4.9	37.2	5.01	7.4
3-A	57.1	8.0	46.1	5.7	41.6	9.8	52.9	2.0	15.5	4.22	3.7
4-B	57.1	6.9	39.4	4.8	39.1	6.2	33.2	2.5	18.0	3.62	5.0
5-A	54.1	5.2	29.6	3.6	24.3	3.4	18.6	3.9	29.8	2.93	10.2
6-B	49.5	6.1	34.8	4.2	30.5	6.0	32.3	2.5	19.0	3.48	5.4
C	90.4	13.2	75.6	9.3	73.3	11.8	63.8	2.2	17.1	5.0	3.4
D	98.0	13.4	76.9	9.5	57.4	15.2	81.8	5.3	40.6	5.35	7.6
C & D	173.6	21.5	123.6	15.2	89.3	23.3	125.3	11.1	84.3	6.69	12.6

Analysis of Resemblance Coefficients

Unfortunately, the wide age range of the subjects tested and the small number of cases at each age level make the interpretation of resemblance coefficients somewhat hazardous. This is especially so in light of the fact that an analysis of the data indicates a curvilinear relationship of test performance to chronological age. For this reason the product-moment partial correlations in Table IX are presented only as a possible suggestion of what the differences in resemblance between fraternal and identical twins might be if age were held constant experimentally.

Despite the limits of interpretation, however, it is interesting to note that the partial correlations tend very definitely to substantiate the findings indicated in an analysis of intra-pair differences. The data at hand, then, can leave very little doubt as to the fact that identical twins resemble one another much more closely in performance on the Spatial Relations Test than do fraternal pairs. A final evaluation of this point, however, must rest on a study of an adequate number of twin pairs of the same age.

Of interest to this study is McNemar's (6) report of the effects of practice on twin resemblances in motor skills. His analysis based on homogeneous groups

Table IX

Resemblance Coefficients in Mechanical Ability*

	Fraternal			Identical		
	Total Correlation	Relationship of Test Performance to Age	Partial Correlation (age held constant)	Total Correlation	Relationship of Test Performance to Age	Partial Correlation (age held constant)
1-A	.625	.578	.438	.748	.537	.646
2-B	.538	.659	.184	.760	.545	.659
3-A	.595	.775	-.015	.708	.498	.613
4-B	.678	.686	.393	.747	.535	.645
5-A	.648	.703	.303	.889	.624	.819
6-B	.674	.676	.399	.823	.462	.775
C	.622	.636	.366	.733	.502	.643
D	.445	.612	.113	.690	.372	.640
C & D	.642	.670	.351	.803	.452	.752

*Resemblance coefficients calculated from double entry scattergrams.

(i.e. with respect to range of scores) showed that practice on the pursuit rotor and spool packing tests increased the resemblance of fraternal twins but that it had no effect on the degree of resemblance of identical twins. On the card sorting test both the fraternal and identical twin pairs showed a slight decrease in resemblance.

Intelligence Test Results

Since it is doubtful that the I.Q.'s on group tests have the same significance at the different age levels, the results on the Kuhlman-Anderson tests must be interpreted with caution.

The mean I.Q. for the total number of twin subjects was 98.35, for the identical group 99.31, and for the fraternal group 97.5. It is interesting to note that the results conform to the findings of Day (3), who in summarizing the literature on the differences in mean I.Q. between fraternal and identical twins, reports that in each of the research studies reported, the identical twins show a consistent superiority to the fraternal twins. The reason for this is not apparent.

The resemblance coefficients in intelligence presented in Table X are lower than those usually reported in the literature. As two typical studies the results of Wingfield (12) and of Merriam (7) are cited. A comparison of the standard deviations in Table X, however, make it evident that the smaller values of r in the present study may be explained by the comparative difference in variability.

Table X

Resemblance Coefficients and Standard Deviations of I.Q.'s on Merriman's, Wingfield's and the Present Studies

	Merriman		Wingfield		Present Study	
	r	S.D.	r	S.D.	r	S.D.
Fraternal	-	-	.70	13.5	.558	9.17
Identical	-	-	.90	14.3	.844	9.86
Combined	.782	15.1	.75	12.65	.705	9.54

Table XI shows that the relationship between performance on the Spatial Relations Test and intelligence (as measured by the Kuhlman-Anderson test) ranges from .031 to .216 on the separate practice trials. This low degree of relationship between mechanical ability and intelligence is in harmony with the test results on 7th and 8th grade children as reported in the Minnesota study (9).

Table XI

Correlations between Time Scores on the Spatial Relations Test and I.Q.'s on the Kuhlman-Anderson Test

Trial	Correlation*
1-A	.216
2-B	.114
3-A	.031
4-B	.162
5-A	.140
6-B	.083
C	.172
D	.134
C & D	.173

* P.E. r for r = o is ±.06.

SUMMARY AND CONCLUSIONS

The purpose of the present study was to determine the degree of twin resemblances in performance on the Minnesota Spatial Relations Test, which is a form board designed to measure mechanical ability. The test was administered to 62 pairs of twin boys of grade school age, 29 of which were diagnosed as identical and 33 as fraternal. The mean I.Q. of the fraternal twins was 97.5 and of the identical twins 99.3. Six successive trials on Boards A and B were given as a practice series. This was followed by the final test on Boards C and D. The results follow:

1. The reliability coefficients for the separate Boards are .853 for A and B and .885 for C and D.
2. The correlations between the practice trials and the combined score on C and D suggest that the same function is being measured on all practice trials.
3. A study of practice effects revealed that (a) variability of scores decreased markedly with practice, (b) the practice curve was negatively accelerated, and (c) there was a transfer effect from the practice to the final series.
4. An analysis based on intra-pair differences and resemblance coefficients showed a markedly closer resemblance between identical twins than between fraternal twins.
5. There is apparently no relationship between twin similarity and the number of practice trials.
6. There is no relationship between performance on the Minnesota Spatial Relations Test and intelligence.

It may be concluded from an analysis of intra-pair differences and from the trends indicated by the resemblance coefficients that identical twins resemble one another much more closely in performance on the Minnesota Spatial Relations Test than do fraternal pairs. The most plausible explanation of this difference in the degree of resemblance between the two groups of twins is that for fraternal pairs there is a greater differentiation of hereditary capacities. It hardly seems likely that the greater similarity of environment between the identical twins would account for the obtained results in that all conditions of testing were held constant for both sets of twins.

The results on twin resemblances in themselves do not allow for a complete evaluation of the environmental factor as it affects mechanical ability. They merely indicate that underlying the ability to perform on the Minnesota Spatial

Relations Test is an important hereditary factor which cannot be neglected. Hogben's (4) point of view is particularly relevant to this aspect of the problem. He states: "The study of twin resemblance shows that the inborn capacities of individual men and women are different. Of itself it does not assist us to form any judgment concerning the extent to which such inborn differences are related to existing inequalities of social privilege...."

BIBLIOGRAPHY

(1) Dahlberg, G.: Twin births and twins from a hereditary point of view. Stockholm: Bokforlags-A,-B. Tidens Trycheri, 1926, pp. 296 + 15 + 18.

(2) Darley, J. G.: The reliability of the tests in the standard battery, in: Research studies in individual diagnosis, edited by D. G. Paterson. Univ. of Minn. Employment Stabilization Research Institute. Bulletin, Vol. III, No. 4, August, 1934.

(3) Day, E. J.: The development of language in twins. II, The development of twins: Their resemblances and differences. Child Development, 1932, 3 298-316.

(4) Hogben, L.: The problem of twin resemblances. Chapter I in: Genetic principles in medicine and social science. New York: Alfred A. Knopf. 1932, pp. 11 + 230.

(5) Jones, H. E. and Wilson, P. T.: Reputation differences in like-sex twins. J. Exper. Educ., 1932-33, 1, 86-91.

(6) McNemar, O.: Twin resemblances in motor skills, and the effects of practice thereon. J. Genet. Psychol., 42, 70-99. 1933.

(7) Merriman, C.: The intellectual resemblance of twins. Psychol. Review Monog., 1924, 33 (No. 152) p. 58.

(8) Newman, H. H.: Studies of human twins. Biol. Bull., 1928, 55, i. Methods of diagnosing monozygotic and dizygotic twins, 283-297. ii. Asymmetry reversed, or mirror imaging in identical twins, 298-315.

(9) Paterson, D. G., Elliot, R. M., Anderson, L. D., Toops, H. A., and Heidbreder, E.: Minnesota Mechanical Ability Studies. Minneapolis: University of Minnesota Press, 1930, pp. xiii + 590.

(10) Seimens, H. W.: The diagnosis of identity in twins. J. Hered., 1927, 18, 201-209.

(11) Stocks, Percy: A biometric investigation of twins and their brothers and sisters. Annals of Eugenics, Univ. Press, Cambridge, Vol. IV, Parts I and II, April 1930, pp. 49-107.

(12) Wingfield, A. H.: Twins and orphans: The inheritance of intelligence. London and Toronto: J. M. Dent & Sons, Ltd. 1928, pp. xii + 127.

THE MENTAL DEVELOPMENT OF CHILDREN OF THE SAME IQ IN DIFFERING INSTITUTIONAL ENVIRONMENTS

ORLO L. CRISSEY[1]

THE PROBLEM

The concept of intelligence as fixed and unmodifiable, or only modifiable within very narrow limits, is yielding to a viewpoint which conceives of intelligence in more functional terms. No longer can it be said that an early intelligence test rigidly classifies a child for life in terms of mental potentialities. Rather, changes in functional mental performance are seen to be related to environmental differences. Recent studies, especially of children at the younger ages, have shown that these changes may be large in extent and appear to follow certain trends in relation to general environmental influences.

Studies of preschool and school age children by Wellman (4) and Skeels (3) have shown that groups of individuals at the mental levels below the central tendency of a population appear to gain on retests, while those at the upper levels tend to gain less or to lose in IQ. The factors responsible for these phenomena are difficult to understand. Do gains (or smaller losses) at the lower IQ levels of a population mean that the average sets the "stimulation level" for the group, tending thus to encourage those at the lower levels to greater attainment, while those in the upper categories are not challenged to attain their relative potentialities to a similar degree? Perhaps Sherman and Key (2) are right when they explain the general decreases in intelligence quotients of children in an isolated Virginia community as due to living in an environment which does not demand greater development. Do children develop only as the environment demands development? Or, applied in another way, does an environment "geared" to the average child of a population tend to stimulate more highly those individuals below the central tendency of the group, by demanding more of them, and fail to stimulate similarly those above the average? Do environments of differing mental levels show differences in these relationships?

If the mental level of an environment is an important factor in a child's mental development, then the rate of mental development of a child in an institution designed for normal and dull-normal children should vary from that of a child of similar mental ability in an institution designed for the feeble-minded. It is the aim of this study to investigate the mental development of children of the same mental level in relation to residence in institutional environments of differing mental level.

SUBJECTS

The subjects for this investigation consist of children resident in four Iowa institutions: the Iowa Soldiers' Orphans' Home at Davenport, the State Juvenile Home at Toledo, the Institution for the Feeble-Minded at Glenwood, and the Hospital for Epileptics and School for Feeble-Minded at Woodward. The cases used were

[1] From Iowa Child Welfare Research Station, State University of Iowa, Iowa City, Iowa.

CHILD DEVELOPMENT, Vol. 8, No. 3 (September, 1937)

selected from a large number of children who had been given individual tests, and who were resident in these institutions between an initial test and one or more retests. Subjects who were over sixteen years of age at the time of the initial test or retest were eliminated so as to make all intelligence quotients computed on the basis of actual life age. No colored children, epileptics, physical anomalies, or cases at the idiot or low imbecile levels were included.

For convenience, the following key will be used in referring to these institutions:

Institution A: The Iowa Soldiers' Orphans' Home at Davenport
Institution B: The State Juvenile Home at Toledo
Institution X: The Institution for Feeble-Minded at Glenwood
Institution Y: Hospital for Epileptics and School for Feeble-Minded at Woodward

TESTS AND PROCEDURE

The Stanford and Kuhlmann revisions of the Binet scale were used for all individual tests. The use of the Kuhlmann, however, was limited to only a few younger children for whom a basal mental age of three years could not be obtained on the Stanford. All tests were administered by well-trained examiners as a part of the co-operative program between the Iowa Child Welfare Research Station and the Iowa Board of Control of State Institutions. These tests have been given at varying intervals during the last eight years.

The approach used in this study is the method of matched groups. In order to obtain as fine control as possible, three criteria were decided upon as the bases for selection in pairing:

1. Individuals must be within 3 points in IQ on initial test.
2. At the time of the first test, chronological age must not vary more than six months.
3. The length of intervals between the respective initial tests and retests must be within six months.

Various environmental comparisons were set up by pairing individuals in homes for dependents with children in institutions for the feeble-minded on these bases. Residents in the homes for dependents were paired not only with nontransfers in the schools for the feeble-minded, but also with children who had been previously transferred from the homes for dependents. In most cases only one individual was paired with another, but in some instances where the criteria were met, one individual might enter into two or more pairings. These cases were the exception rather than the rule. The following environmental groupings are compared in relation to mental development.

1. Residents at homes for dependent children, Institutions A and B, with nontransfers in schools for the feeble-minded, Institutions X and Y
2. Residents at Institutions A and B with transfers from these institutions to Institution X
3. Residents at Institution A with transfers from Institutions

A and B to Institution Y

4. Residents at Institution B with transfers from A and B to Y
5. Total pairings of residents at Institutions A and B with transfers and nontransfers at Institutions X and Y by chronological age groupings (under six, seven to twelve, thirteen and over)
6. Total pairings of residents at Institutions A and B with transfers and nontransfers at Institutions X and Y

RESULTS AND CONCLUSIONS

Analyses of these comparisons are presented below in two tabulations. In the first, children of the same IQ resident in environments of differing mental level are studied on the bases of the various environmental groupings. In the second tabulation these pairings have been classified according to age groupings. The results of these analyses are as follows:

Matched Groups	Children	Net Changes in IQ			IQ	
		Mean	S.D.	$S.E._M$	Mean	Range
Residents at Institutions A and B	48	.5	6.0	.9	68.6	51 to 98
Residents at Institutions X and Y	48	-5.5	7.1	1.0	68.9	52 to 99
Residents at Institutions A and B	30	3.6	7.4	1.4	69.0	60 to 74
Transfers to Institution X	30	-4.7	7.3	1.3	68.9	61 to 73
Residents at Institution A	49	2.6	7.8	1.1	68.4	50 to 75
Transfers to Institution Y	49	-3.9	9.9	1.4	68.3	49 to 77
Residents at Institution B	39	1.7	5.7	.9	70.4	65 to 73
Transfers to Institution Y	39	-4.5	3.7	.6	70.1	63 to 73

Age Groupings	Children	Net Changes in IQ			IQ	
		Mean	S.D.	$S.E._M$	Mean	Range
Residents in Institutions A and B						
Under six years	21	8.5	10.2	2.2	64.1	50 to 82
Seven to twelve years	76	.2	6.2	.7	70.3	56 to 98
Thirteen years and over	50	2.9	4.6	.7	68.2	59 to 73
Total all ages	147	2.3	7.0	.6	68.7	50 to 98
Residents and Transfers in Institutions X and Y						
Under six years	21	-4.3	11.8	2.6	64.3	49 to 81
Seven to twelve years	76	-5.5	5.6	.6	70.3	55 to 99
Thirteen years and over	50	-4.2	3.3	.5	68.4	62 to 76
Total all ages	147	-4.9	7.8	.6	68.8	49 to 99

They may be summarized in the statements below:

1. When children at Institutions A and B are paired with nontransfers at Institutions X and Y, the orphanage children tend to remain constant while the

children in the institution for the feeble-minded tend to lose. A significant difference of 6.0 IQ points is found between these paired groups. (Critical ratio, 4.5)

2. Comparison of residents of the two homes for dependents (A and B) and transfers from these homes to an institution for feeble-minded (X) shows a difference of 8.3 points, and is statistically significant. (Critical ratio, 4.4)

3. Residents at Institution A when paired with transfers from Institutions A and B to Institution Y indicate a significant difference (6.6 points) in favor of the orphanage group. (Critical ratio, 3.6)

4. When children at Institution B are paired with children transferred from the two homes for dependents to Institution Y, a similar significant difference may be noted. (Difference, 6.2 points; critical ratio, 5.7)

5. Regardless of age, the orphanage children continue to stay constant or gain while their pairs in schools for the feeble-minded tend to lose. In each case significant differences were obtained. (Differences, 12.8, 5.8, 7.1; respective critical ratios, 3.7, 6.1, 8.8)

6. Throwing all individuals from the orphanages who entered into pairings into one group, and similarly combining all the paired individuals in the schools for the feeble-minded, a total of 147 pairs is obtained. In such a composite some pairs doubtless are entered several times. Their number is not large enough, however, to influence to any great extent the total effect. The results show a reliable mean gain of 2.3 IQ points for the orphanage children, while their pairs in schools for the feeble-minded show a mean loss of -4.9. The difference is significant. (Difference, 7.2; critical ratio, 8.5)

These findings reveal a high degree of internal consistency. In every comparison there is a consistently significant difference between these carefully paired individuals in environments which differ at least 30 IQ points in mental level. In not a single comparison do the means of the orphanage groups show a loss, while a loss is indicated in every mean for the groups in the institutions for the feeble-minded. It should be remembered that in the homes for dependents these children represent the lower mental levels of the institutional population while in the schools for the feeble-minded those of similar IQ are in the higher strata. Seemingly, institutional environments of differing mental levels present unlike demands upon children of the same IQ, causing variations in the rate of mental development in accordance with the child's relative placement above or below the mental level of his environment. In other words it would appear that the average tends to set the "stimulation level" for the group, and children develop as the environment demands development. This conclusion is further substantiated in other findings in the larger study of which this investigation is a part (1).

REFERENCES

(1) Crissey, Orlo L.: Mental development as related to institutional residence and educational achievement. Univ. Iowa Stud., Stud. in Child Welfare, 1937, 13, No. 1, Pp. 81.

(2) Sherman, Mandel, and Key, Cora B.: The intelligence of isolated mountain children. Child Development, 1932, 3, 279-290.

(3) Skeels, Harold M.: A cooperative orphanage research. J. Educ. Res., 1937, 30, 437-444.

(4) Wellman, Beth L.: The effect of pre-school attendance upon the IQ. J. Exper. Educ., 1932-1933, 1, 48-69.

VALIDITY OF COMPARISONS OF CURRENT METHODS FOR ESTIMATING PHYSICAL STATUS

RAYMOND FRANZEN

A paper by Everett L. Marshall in the March 1937 number of "Child Development," prompts a brief discussion of the basic considerations which should control an evaluation of physical status estimates.[1]

There are three ways in which the discussion of this subject becomes confused. The paper used for purposes of illustration in this criticism is guilty of all of these. The departures from clear analysis are:

1. Failure to distinguish the severity of selection from the adequacy of the selection. Any method of measurement may be set to obtain any number of selections desired, but the main criterion for its value is the degree to which the selections which are made, represent what is claimed for them.
2. The assumption that "under-weight" is synonymous with "under-nourished."
3. The reference to popular abbreviated applications without due consideration of the basic published material which is their foundation.

DISTINCTION BETWEEN SEVERITY OF SELECTION AND ADEQUACY OF SELECTION

One of the comparisons made in the paper here used as illustration is between three methods of selection which use progressively more skeletal measures as control. The Baldwin-Wood tables are based on a height-weight correlation. The Pryor and Stoltz tables are based on the correlation of height and hips with weight.[2] The third method used in comparison is that of C. H. McCloy which uses weight residuals from a still more inclusive skeletal arrangement.

It is statistically obvious that the variability of the residuals in weight must grow progressively smaller as the skeletal elements in the correlated variable increase. This is best shown by comparison of the multiples and residuals involved as presented in the 1929 American Child Health Association Monograph on this subject.[3]

The real problem is not the shrinkage in variability (and this is all that is shown in Table I of Marshall's paper) but rather an evaluation by outside criteria of the cases which are part of a critical group (selected by standard deviation position or by percental methods). Studies have been made which arrive at such an evaluation and it can be shown that children selected by use of multiples containing skeletal measurements in addition to height are far more likely to be the type which physicians call undernourished than are those selected by either the Baldwin-Wood tables or the Pryor and Stoltz tables. From the point of view of

[1] Everett L. Marshall. "A Comparison of Four Current Methods of Estimating Physical Status."

[2] Pryor and Stoltz do not use the multiple but do derive a type of relationship. The multiple gives better results.

[3] Franzen, Raymond. "Physical Measures of Growth and Nutrition." Monograph II of the School Health Research Series. American Child Health Association, New York, 1929.

method, there is no distinction in this better type of weight selection between the one published by C. H. McCloy in 1936 and 1937 and the American Child Health Association method published in 1929, "Physical Measures of Growth and Nutrition."

The comparisons given by Marshall in Table II do attempt to show the overlap of selection but the error of unequal severity of selection makes conclusions from this table impossible. In order adequately to compare the methods, the measurements must be so set that each selects the same number of critical cases. This follows because the function we are trying to measure is a variable and not a category and any attempt to compare selections made by any two methods must insure the same rigor for each. One could obviously set a particular measurement to include 90% of the cases and then this would increase one's chances of overlap with other methods of selection enormously.

CONFUSION OF "UNDERWEIGHT" WITH "UNDERNOURISHED"

Marshall in his paper continuously refers to the ACH [1] as an estimate of "being underweight." This was, of course, never intended and should not be imputed as an objective. The estimate was purposely built to avoid the use of weight in the consideration of malnutrition. The reasons for this procedure are given in the monograph referred to above, which deals wholly with the matter of estimate of weight by multiples. The authors have produced tables for use in making weight estimates as residuals in a multiple regression and these are entirely independent of the ACH index.

The value of this index can in no sense be derived from a comparison with weight selections. An attempt has been made to evaluate selections in terms of a more inclusive anthropometric survey than is involved in either of the three methods quoted in the article under discussion.[2] This paper showed the degree of success by this type of estimate when compared with estimates of comparable severity in two of the other measurements under discussion. The objective was to find children who showed an overlapping deficiency in three criteria, - weight for skeletal build, arm girth for skeletal build, and subcutaneous tissue for skeletal build. When these three criteria are used in combination the ACH screen selection is far superior to that made by the two other methods. It netted 41 out of 43 (as yielded by an intensive survey of 500). This compares to 27 and 16 out of 43 netted by height-hip-weight and height-weight methods respectively.

More complete discussion of the adequacy of the use of the ACH for the selection of nutritional defects is contained in Chapter 7 of "Physical Defects, The Pathway to Correction," American Child Health Association 1934.

[1] Franzen, R. and Palmer, G. T. "The ACH Index of Nutritional status." New York American Child Health Association, 1934.

[2] Franzen, Raymond. "Selection of Malnourished School Children." American Journal of Diseases of Children, April 1934, Vol. 47, pp. 789-798.

CORRELATIONS OF PERCEPTION WITH OTHER ABILITIES AND TRAITS IN GRADE I[1]

FRANK T. WILSON AND CECILE WHITE FLEMMING

I. INTRODUCTION

During the school year 1933-34 a variety of tests was given to twenty-five children in Grade I of the Horace Mann School, Teachers College. These included tests of "reading readiness"; many of the Gates Reading Diagnosis tests; some reading achievement tests; mental ability tests, such as the Stanford Revision of the Binet-Simon tests and various performance tests; certain psychological tests, as of perception and perseveration; and several measures of psycho-physical and personality traits and of home background. The purpose of the study was to examine any possible relationships that might exist between measurable traits and abilities, and early progress in the mechanics of reading.

The children of the group came from well-to-do homes. A large percentage of the parents were professional people. The following averages for these pupils were found.

Chronological Age	6.31
Mental Age	7.61
Intelligence Quotient	120.6

Nearly every test and measurement was given or made individually, under carefully controlled conditions, and by reliable persons accustomed to administering tests to young children. The cooperation of the pupils was almost invariably excellent. It is believed for these reasons, that errors of examination were unusually low.

This report presents correlations of a perception test with about seventy other measures and appraisals used in the original study. An original test was devised in which the Graflex Vocal Plan Shutter Apparatus, (Stoelting Company) was used. The aperture was set at 1½, and the tension at G-1. The adjustment plate indicated that this gave a reading of 10.

Thirty-two test cards, 6½ by 4½ inches, of white, unglazed cardboard were used for the stimulus symbols. The symbols were: horizontal lines; vertical lines; fowl silhouettes; heavy outlined triangles; the digits 2, 3, 4, 5, and 6; and the capital letters A, B, C, D, and E. The letters and numbers were in random order on their respective cards. There were four exposure cards each for the lines, triangles and silhouettes, and eight for both the numbers and the letters. For each kind of symbol, one card contained 2 symbols, one 3, one 4, and one 5, except that for both the numbers and the letters there were two cards with 2, 3, 4, or 5 symbols. The lines, letters, and figures were made in black India ink with a 1/8 inch wide lettering pen, they were about 5/8 of an inch in altitude, and were spaced about 1/2 to 5/8 of an inch apart. At the beginning of the experiment, the

[1] This report presents a minor phase of a study of Reading Readiness and Reading Progress in the Primary Grades of the Horace Mann School, Teachers College, New York, 1933-36. This study has been made possible by the cooperation of Miss Agnes Burke, Teacher of Grade I and other teachers of Kindergarten and Primary Grades. It has been made under the supervision of Doctor Cecile White Flemming, Director of Pupil Individual Development and Guidance, and of Doctor Rollo G. Reynolds, Principal. Prepared by F. T. Wilson, Department of Education, Hunter College, New York, with the assistance of the U. S. Works Progress Administration, New York City, project number 65-97-295, sub-project 25.

cards were shuffled and during the test were presented to every child in the order thus obtained. Two practice cards were used before the test was begun in order to familiarize the children with the purpose and method of the experiment. The first practice card had two fowls; the second, the letters X, A, and J. The description and order of the cards are shown in Table 1.

Table 1

Decription of Perception Test Cards

				Order of Presentation of Test Cards
Practice	Card	a.	2 ducks (in silhouette)	
"	"	b.	X A J	
Test	"	1.	2 horizontal lines.....	31
"	"	2.	3 " "	28
"	"	3.	4 " "	19
"	"	4.	5 " "	11
"	"	5.	2 vertical lines	7
"	"	6.	3 " "	2
"	"	7.	4 " "	8
"	"	8.	5 " "	10
"	"	9.	2 hens (in silhouette)	16
"	"	10.	3 " (" ")	3
"	"	11.	3 roosters and 1 duck (in silhouette)..	29
"	"	12.	5 roosters (in silhouette)	17
"	"	13.	65	32
"	"	14.	462	6
"	"	15.	4362	25
"	"	16.	35642	9
"	"	17.	42	26
"	"	18.	243	15
"	"	19.	3246	23
"	"	20.	56243	14
"	"	21.	C E	5
"	"	22.	C E B	1
"	"	23.	A D B C	12
"	"	24.	E B A C D	20
"	"	25.	E D	18
"	"	26.	A E B	27
"	"	27.	C A E B	21
"	"	28.	C B D E A	4
"	"	29.	2 triangles (in heavy outline)	13
"	"	30.	3 " " " "	22
"	"	31.	4 " " " "	24
"	"	32.	5 " " " "	30

The child was seated in a comfortable chair with his back to the window. The apparatus stood on a table before him, the face slanting at an angle of about eighty-five degrees from the horizontal plane. The distance from the child's eyes to the aperture was about fourteen inches. The shutter was operated by a small hand lever trip. The cards were changed between exposures and, of course, were concealed from the child's view except during the instant of exposure. An assistant made the record of responses which were given orally in answer to the questions: "What do you see?" "How many?"

The score was the number of correct responses. In practically every case the symbols were correctly described, but the children varied in reporting the number of symbols perceived. The test seemed quite long for the children tested. A retest of nine children with half of the cards showed an agreement of 58% of responses with the first test for the same sixteen cards. Self-correlation of split halves for the entire test was .55.

The data of the study are in terms of correlations obtained by the rank order method. To secure the rank orders all measures and appraisals were reduced to numerical scores. Owing to lack of facilities, it was not feasible to make all the computations that were possible in the original study. A "finder" device was

used to select for computation the correlations which seemed to promise significance. It is believed that through the use of this device, although it was not altogether accurate, all the high and fairly high correlations were found. The correlations omitted were probably below .50, and most of them probably nearer zero than .50. The P.E. of rho's when N = 25, range from ± .0237 for .90 to ± .1335 for .10.

The validity of many of the measures and appraisals is uncertain. Few correlations of seemingly unusual size were obtained, however, and few which were inconsistent with other correlations for the same kind of traits and abilities found in the complete data of the original study. The opinion of the teacher, of the school psychologist, and of other qualified persons who have studied the figures, is that the results have quite high validity.

II. FINDINGS

List 1 shows all the computed correlations of perception with the other tests in order of size. None of these correlations is very high. Those for Metropolitan similarities and for total matching, .47 and .45 respectively, are about what would be expected from the nature of the abilities involved in the tests. The similarities test required visual perception of similarities and differences. The matching tests required much the same sort of visual perception.

The correlation -.46 with the Gates Diagnosis Test, XI, 2, plus supplement, which required the recognition of two spoken words as being the same or different words, does not seem to be explicable on any grounds but chance. This test was auditory. The perception test was visual. That there should be fairly high negative correlation between these differing abilities is hardly probable.

List 1

Correlations of Perception with 73 Other Measures

Metropolitan Readiness Test, Similarities	.47
Total matching	.45
Hildreth First Grade Reading Analysis Test, Matching phrases	.42
Hildreth First Grade Reading Analysis Test, Matching sentences	.38
Gates Reading Diagnosis Tests, XV, 1, Memory span, digits	.37
Grip	.37
Gates Reading Diagnosis Tests, XV, 3, Memory span, nonsense syllables	.36
Seguin, Form Board Time	.36
Manikin, Pintner-Patterson, Time	.35
Metropolitan Readiness Test, Copying	.34
Metropolitan Readiness Test, Information	.34
Stone & Grover Classification Test for Beginners in Reading II	.33
Healy Picture Completion Test, II	.33
Van Wagenen Reading Readiness Test, Word learning	.32
Gates Reading Diagnosis Tests, XV, Total memory span	.31
Nutrition (variation from height-weight-age norms)	.31
Reversals, Auditory perception	.31
Hildreth First Grade Reading Analysis Test, Total	.30
Weight	.30
Gates Reading Diagnosis Tests, XV, 4 Memory span words	.29
Stone & Grover, Classification Test for Beginners in Reading, Total	.28
Teacher's Ranking in Reading, November prediction	.28
Gates Reading Diagnosis Tests, IX, 1-7 & 9. Phonic combinations and letter sounds	.28
Undesirable behavior traits	.27
Gates Reading Diagnosis Tests, IX, 9, Giving letter sounds	.27
Motor Coordination (Battery of six tests)	.26
Hildreth First Grade Reading Analysis Test, Matching words and phrases	.26
Teacher's Ranking in Reading, May	.24
Gates Reading Diagnosis Tests, IX, 1-7, Giving phonic combinations	.23
Gates Reading Diagnosis Tests, VIII, 3, Word recognition, auditory perception	.23
Gates Primary Reading Tests, Type 1, Word recognition, March	.22
Gates Primary Reading Tests, Type 2, Sentence reading, May	.22
Gates Primary Reading Tests, Type 2, Sentence reading, March	.21
Metropolitan Readiness Test, Total	.21

List 1 - Continued

Correlations of Perception with 73 Other Measures

Van Wagenen Reading Readiness Test, Memory span	.21
Van Wagenen Reading Readiness Test, Information	.20
Hildreth First Grade Reading Analysis Test, Matching words	.20
Gates Reading Diagnosis Tests, XV, 2, Memory span, letters	.19
Perseveration, Elkins-Maller, Attention Test	.19
Mare & Foal, Time	.18
Gates Reading Diagnosis Tests, XIII, 3, Adapted, Writing capital and small letters and digits	.17
Tapping, Whipple and Healy	.16
Van Wagenen Reading Readiness Test, Relations	.14
Gates Primary Reading Tests, Type 1, Word recognition, May	.14
Gates Primary Reading Tests, Type 3, Paragraph reading, May	.12
Metropolitan Readiness Test, Drawing man	.11
Gates Reading Diagnosis Tests, XIII, 1-2, Write words	.11
Ship Test	.09
Chronological Age	.08
Nervousness Index	.07
Developmental Index (babyhood)	.06
Van Wagenen Reading Readiness Test, Vocabulary	.05
Personal Traits	.05
Gates Reading Diagnosis Tests, X, 1, Blend letters	.11
Gates Reading Diagnosis Tests, IX, 10, Recognition capital letters	.05
Reversals, Visual perception	.04
Van Wagenen Reading Readiness Test, Word discrimination	.03
Total, Vocabulary Time	.03
Gates Primary Reading Tests, Type 3, Paragraph reading, March	.03
Metropolitan Readiness Test, Vocabulary	.03
Mental Age, Stanford Revision of the Binet-Simon	.02
Stone & Grover, Classification Test for Beginners in Reading I	.02
Gates Primary Reading Tests, Type 1, Word recognition, November	.02
Gates Reading Diagnosis Tests, VIII, 2, Word recognition, visual presentation	.01
Metropolitan Readiness Test, Numbers	-.001
Total Vocabulary	-.004
Perseveration, Elkins-Maller, Attention Test	-.03
Total Errors	-.03
Gates Reading Diagnosis Tests, IX, 11, Recognition small letters	-.04
Parent Questionnaire, No. of Activities, Total	-.04
Intelligence Quotient, Stanford Revision of the Binet-Simon	-.07
Steadiness (hole apparatus)	-.08
Gates Reading Diagnosis Tests, X, 2, Recognition sounded letters	-.08
Mare & Foal, Errors	-.08
Personality Rating, Hicks, A Personality Rating Scale for Children Six to Nine	-.09
Height	-.09
Gates Reading Diagnosis Tests, XI, 2, Same-different words	-.12
Writing Time	-.14
Metropolitan Readiness Test, Sentences	-.15
Gates Reading Diagnosis Tests, X, Total, Auditory perception	-.18
Gates Reading Diagnosis Tests, XI, 1, Repeat nonsense syllables	-.20
Gates Reading Diagnosis Tests, X, 3-4, Giving initial and final sounds	-.21
Gates Reading Diagnosis Tests, XI, 2 and supplement, Same-different words	-.46
Average	.138

Lists 2 to 6 give correlations of perception with groupings of other measures. List 2 shows the correlations of perception with fourteen reading tests. Only two of these correlations are even fairly high, namely, .42 and .38. Both are for subtests of the Hildreth First Grade Reading Analysis Test, one requiring matching phrases and the other matching of sentences. The total of this test had a correlation of .30 with perception. The teacher's rankings in reading for November and May gave correlations of .28 and .24, respectively. None of the reading tests had negative correlations with perception, although the two lowest ones were .02 and .03. The average for the fourteen correlations was .217, and the range was from .02 to .42. Some but not very close relationship between the perception and reading tests is indicated.

List 3 gives the correlations between the perception test and tests of abilities with letters. There are five negative correlations in this list. All of these are for letter tests involving sounds and requiring no use of the eyes. One other test in this group of abilities with letters, a memory span test, was

List 2

Correlations of Perception with 14 Reading Tests

Gates Primary Reading Tests, Type 1, Word recognition, November	.02
" " " " " 1, Word recognition, March	.22
" " " " " 2, Sentence reading, March	.21
" " " " " 3, Paragraph reading, March	.03
Hildreth First Grade Reading Analysis Test, Matching phrases	.42
" " " " " " Matching words	.20
" " " " " " Matching sentences	.38
" " " " " " Matching words and phrases in sentences	.26
Total " " " " "	.30
Teacher's ranking in reading, November prediction	.28
" " " " May, ability	.24
Gates Primary Reading Tests, Type 1, Word recognition, May	.14
" " " " " 2, Sentence reading, May	.22
" " " " " 3, Paragraph reading, May	.12
Average	.217
Range .02 to .42	

List 3

Correlations of Perception with 17 Tests of Letter Abilities

Stone & Grover, Classification Test for Beginners in Reading I	.02
" " " " " " " " II	.33
Gates Reading Diagnosis Tests, VIII, 2, Recognition of words seen	.01
" " " " VIII, 3, " " " " heard	.23
" " " " IX, 1-7 Phonic combinations	.23
" " " " IX, 9, Giving letter sounds	.27
" " " " IX, Total 1-7 and 9	.28
" " " " X, 1, Blend words	.11
" " " " X, 2, Recognition sounded letters	-.05
" " " " X, 3-4, Giving initial and final sounds	-.21
" " " " X, Total auditory perception	-.18
" " " " XIII, 1-2, Writing words	.11
" " " " XIII, 3, Writing letters	.17
" " " " IX, 10, Recognition capital letters	.05
" " " " XI, 2, Recognition same-different words	-.12
" " " " XI, 2, plus supplement	-.46
" " " " XV, 2, Memory span, letters	.19
Average	.058
Range -.46 to .33	

List 4

Correlations of Perception with 20 Tests of Mental Ability

Van Wagenen Reading Readiness Test, Information	.20
" " " " " Relations	.14
" " " " " Vocabulary	.05
" " " " " Memory span	.21
Metropolitan Readiness Tests, Vocabulary	.03
" " " Sentences	-.15
" " " Numbers	00
" " " Information	.34
" " " Drawing Man	.11
Gates Reading Diagnosis Tests, XI, 1, Repeat nonsense syllables	-.20
Vocabulary, Time (Binet, Action-Agent, Kindergarten Word List)	00
Seguin Form Board	.36
Mare & Foal, Time	.18
Mare & Foal, Errors	-.08
Manikin	.35
Ship Test	.09
Healy Picture Completion Test, II	.33
Mental Age (Binet)	.02
Intelligence Quotient (Binet)	-.07
Gates Reading Diagnosis Tests, XV, 1-4 Total memory span	.31
Average	.078
Range -.20 to .36	

List 5

Correlations of Perception with 14 Psycho-Physical Measures

Total Writing Time	-.14
" Vocabulary Time (Binet, Action-Agent, Kindergarten Word List	.03
Steadiness (hole test)	-.08
Tapping Whipple-Healy	.16
Perseveration, Elkins-Maller, Attention Test, Second test	.19
Chronological Age	.08
Grip	.37
Perseveration Elkins-Maller Attention Test, First test	-.03
Motor Coordination (6 tests)	.26
Weight	.30
Height	.09
Nutrition	.31
Total Number Activities, parent questionnaire	-.04
Developmental Index (Babyhood)	.06
Average	.099
Range -.14 to .37	

List 6

Correlations of Perception with 7 Personality Measures

Reversals, Visual perception	.04
" Auditory perception	.31
Total Errors (9 tests)	-.03
Undesirable behavior traits, parent questionnaire	.27
Personal Traits	.06
Personality Rating, Hicks, A Personality Rating Scale for Children Six to Nine	-.09
Nervousness Index, parent questionnaire	.07
Average	.089
Range -.09 to .31	

also non-visual. It correlated .19 with the perception test. Of the twelve visual tests, two correlated .02 and .01 with the perception test. They were Part I of the Stone and Grover Classification Test and the Gates Reading Diagnosis subtest for recognition of words seen. In both of these tests the children were to compare printed words. The average for all of the seventeen correlations of the list was .056, and the range was from -.46 to .33. Quite varied, but in the main not close, relationships between the perception test and the letter ability measures is indicated by these figures.

The twenty correlations of perception with tests of mental ability, shown in List 4, ranged from -.20 to .36 and averaged .078. Three performance tests, namely the Seguin Form Board, Manikin, and Healy Picture Completion Test II, correlated .36, .35 and .33, respectively. M. A. correlated .02 and I. Q. -.07. None of the correlations seems of special importance.

The correlations of fourteen psycho-physical measures with perception, given in List 5, ranged from -.14 to .37 and averaged .099. The highest, .37 was for grip. C. A. correlated .08 and a composite score for developmental index (in babyhood), .06. There seems to be little evidence of significant relationships between the perception test and the abilities and traits in this group of measures.

List 6 gives correlations with seven personality measures. None of these is high enough to suggest any relationship of consequence.

SUMMARY AND CONCLUSION

A test of perception by an exposure apparatus, with twenty-five Grade I chil-

dren, yielded a few fairly high correlations with some visual matching tests. The average of correlations of the perception test with 14 reading tests was .217; with 17 letter tests, .056; with 20 mental tests, .078; with 14 psychophysical measures, .099; and with 7 personality measures, .089.

These results indicate very moderate degree of relationship between perception, as measured, and tested reading abilities, and varied degrees of relationship, ranging from moderate to little or none, between perception, as measured, and tested abilities with letters, and with measured and estimated mental, psychophysical, and personality traits.

****

CORRELATIONS OF MEMORY SPAN WITH OTHER ABILITIES AND TRAITS IN GRADE I

This report presents correlations derived from six measures of memory span with seventy-six other measures and appraisals of the pupils. The Memory Span Tests were:

1. A sub-test of the Van Wagenen Reading Readiness Tests called "Memory Span for Ideas." This consisted of twenty-five sentences of increasing length. The first sentence was: "Children like to play with dogs and cats"; the second: "He may get new shoes for Christmas"; and the others were of like nature.

2. The Gates Diagnosis Reading Tests, XV, Memory Span, consisting of four parts: digits, letters, nonsense syllables and familiar words.

3. Gates Diagnosis Reading Tests, sub-test XI, 1, Repeating Nonsense Words. This test was given as a memory span test of syllables, not as a test of nonsense words, which is the purpose of the test according to the directions. The syllables were successions of miscellaneous short familiar words and short meaningless syllables, as for example, "ho-min-us" "du-ran-gib-a-lin," etc. The number of syllables increases from two to six as progress is made thru the test. This test will be called the mixed syllables test in the report.

The test scores used in the study were:

1. For the Van Wagenen Test, the score value shown in the test directions.

2. For the Gates Diagnosis Tests, the length of the memory span in each part of the test; and a total score made by summing the scores of the four parts of the tests.

The data of the study are in terms of averages and correlations obtained by the rank order method. To secure the rank orders all measures and appraisals were reduced to numerical scores. Owing to lack of facilities, it was not feasible to make all the computations that were possible in the original study. A "finder" device was used to select for computation the correlations which seemed to promise significance. It is believed that through the use of this device, although it was not altogether accurate, all the high and fairly high correlations were found. The correlations omitted were probably below .50, and most of them probably nearer zero than .50. Representative P. E.'s of rho's, when N = 25, are given in Table 2.

The validity of many of the measures and appraisals is uncertain. Few correlations of seemingly unusual size were obtained, however, and few which were inconsistent with other correlations for the same kind of traits and abilities as found in the complete data of the original study. The unanimous opinion of the teacher, of the school psychologist, and of other qualified persons who have studied the figures, is that the results have quite high validity.

II. FINDINGS

A - INTERCORRELATIONS

Table 1 shows the intercorrelations of the six memory span tests with each other. The average of all of the fifteen coefficients of intercorrelations of the subtests is .53. The highest intercorrelations are, naturally, those of the Gates total memory span test score with the several parts that make up that total score. They range from .71 ± .064 to .88 ± .0283. The intercorrelations of the separate tests ranged from .29 ± .122, to .71 ± .064. The latter was the coefficient for the letters and digits tests. The large P. E.'s and the varied sizes of the intercorrelations indicate lack of high reliability for the measures. The intercorrelations of the two tests using senseless syllables, namely the test for nonsense syllables and that of mixed short words and meaningless syllables, gave

Table 1

Intercorrelations of Memory Span Tests

		Gates Reading Diagnosis Tests						
	Van Wagenen Sentences	Digits	Letters	Nonsense Syllables	Fam. Words	Mixed Syll.	Gates Total* Col.2-5	Ave. Col.1-6
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Van Wagenen, Sentence Memory Span	--	.59	.42	.42	.49	.68	.69	.52
Gates, Digits	.59	--	.71	.57	.63	.49	.88	.598
Gates, Letters	.42	.71	--	.54	.52	.43	.81	.524
Gates, Nonsense Syllables	.42	.57	.54	--	.60	.29	.79	.484
Gates, Familiar Words	.49	.63	.52	.60	--	.55	.71	.558
Mixed Syllables	.68	.49	.53	.29	.55	--	.43	.49
Average	.52	.598	.524	.484	.558	.49		
*Gates Total, Digits, Letters, Nonsense Syllables, Familiar Words	.54	.88	.81	.79	.71	.43	.693	

Frequency Distribution of Table 1, Excluding Averages and Totals

.70 to .79	1
.60 to .69	3
.50 to .59	5
.40 to .49	5
.30 to .39	0
.20 to .29	1
	15

Average = .53 S. D. = ± .117

Table 2

P. E. of Representative RHO Correlations when N = 25 are as Follows:

Rho		P. E.	Rho		P. E.
.90	±	.025	.55	±	.091
.85	±	.035	.50	±	.099
.80	±	.046	.40	±	.112
.75	±	.056	.30	±	.122
.70	±	.066	.20	±	.129
.65	±	.075	.10	±	.134
.60	±	.083			

the rather surprisingly low figure .29. The remaining intercorrelations of the mixed syllables test with the other tests ranged from .43 to .68, the highest being with the sentence span. The coefficient for mixed syllables test and the total of the Gates tests was .43. The average of the intercorrelations of this test with the other tests, .49, is not appreciably different from the averages of the other tests.

It might be supposed that close similarity in test materials would have produced higher intercorrelations than tests of dissimilar materials. For example, since sentences are made up of words perhaps the intercorrelation between the sentence test and the familiar word test would be higher than most of the other intercorrelations. This coefficient is .49, however, as compared with .59 for the correlation of sentences and digits, and .68 for that of sentences and mixed syllables. Again, words are composed of letters, but that similarity failed noticeably to affect the intercorrelation between the tests using words and letters, as the coefficient, .52, is practically the same as the average of the entire fifteen intercorrelations, which was .53. The highest intercorrelation, .71, was for letters and digits. Possibly the similarity in form - short and simple - rather than similarity in content made that relationship somewhat closer. The wide variation in the figures of the table seems to indicate that the abilities concerned were not simple or unitary abilities but complex ones, probably involving many factors which varied greatly from one test situation to another.

B - AVERAGES

Averages, standard deviations, and probable errors of the averages of the several memory span tests are shown in Table 3 for two groups.

Table 3

Average Scores of Memory Span Tests Grade I Pupils, Groups A and F

Groups	Average		S. D.		P.E.	Average
	A	F	A	F	A	F
Time Tested	Fall	Spring				
Number Pupils	25	30				
Van Wagenen						
Sentences	44.70		11.77		1.59	
Gates Diagnosis						
Digits	5.90		.90		.132	
Letters	4.98	5.31	.854	.895	.115	.108
Nonsense Syllables	4.66		.784		.106	
Familiar Words	4.78	4.86	.67	.542	.086	.066
Mixed Syllables	5.26		.861		.116	
Total*	18.86		2.726		.386	

*Total Digits, Letters, Familiar words, and Nonsense Syllables tests of the Gates Reading Diagnosis Tests.

All of the terms are given for Group A, which was the grade studied intensively in this investigation. They are given only for the letter and word tests for the other group, Group F. Comparison of the averages made by Group A, in the Gates sub-tests shows that the averages for the digits test was highest, 5.90. The average for the mixed syllables test was next highest, 5.26; then that for letters, 4.98; next that for familiar words, 4.78; and last that for nonsense syllables 4.66. It will be noted that the average spans for group F were very nearly the same as those for Group A. The letter span of this group was .33 of a unit longer than the letter span for Group A; and the familiar word span was .12 of a unit longer. When allowance is made for the later date at which the F group was tested

the differences seem to be of little or no significance.

The standard deviations were quite large for all tests, varying from ±.64 to ±.90 for the Gates subtests. The P. E.'s of the averages for these subtests were also quite large, ±.086 to ±.132. It is interesting to note that the largest S. D. and P. E. of the average in this group of tests was for the digits test - which gave the highest average and which seemed, according to the table of intercorrelations, to be the most reliable subtest. On the other hand the smallest S. D. and P. E. of the average were for the familiar word test. This test gave next to the lowest average in the Gates group. It also gave average intercorrelations (Table 1) with all the other tests that were almost as high as those given by the digits test. Table 3 shows that, for these children, the variability in the abilities involved in the tests was very large.

Table 4 gives the critical ratios for the differences between the averages of the four subtests of the Gates tests. The table shows that only the differences

Table 4

The Critical Ratio's of the Differences Between the Averages of the Gates Memory Span Subtests, Grade I Pupils

	Group	Differences between Averages	Critical ratios
Digits less Letters	A	.92	3.54
" " Nonsense Syllables	A	1.24	4.94
" " Familiar Words	A	1.12	4.73
" " Mixed Syllables	A	.64	2.45
Letters less Nonsense Syllables	A	.32	1.38
" " Familiar Words	A	.20	.92
" " " "	F	.45	2.39
" " Mixed Syllables	A	-.28	1.15
Nonsense Syllables less Familiar Words	A	-.12	0.58
Nonsense Syllables less Mixed Syllables	A	-.60	2.58
Familiar Words less Mixed Syllables	A	-.48	2.20

between the digits test and each of the letters, nonsense syllables and familiar word tests are statistically significant. A possible explanation for this fact is that digits may be easier to keep in immediate memory than the other test material. Their written form is simple and the associations of that form with the verbal form may be closer. Perhaps memory images aided the children in holding the spans in mind. There are but nine common visual forms for the nine digits, while there are at least fifty-two common printed plus additional script forms for the twenty six letters. Memory images of words and syllables would probably be still less helpful that those of letters to young children.

The score for the Van Wagenen Memory Span test for sentences was derived from the count of the number of sentences correctly repeated by the subject. The average score for the group, 44.7, corresponds to a total of fifteen sentences repeated correctly. The sentences of the test are numbered consecutively. Fifteen correct does not mean, necessarily, that the sentence numbered fifteen was the longest one which, on the average, the children could repeat. It does seem reasonable, however, to take that sentence as the measure of average ability of the group. This sentence is, "When it storms hard I have to stay in the house to keep dry." This is fourteen syllables long. Van Wagenen entitled this test, "Memory Span for Ideas," but gave no count in either the test or manual of the number of ideas in any of the sentences. The inserted check marks indicate what seem obviously to be five main ideas. Sentence 1 and some others in the test

seem to contain four main ideas. Most of the remainder up to and including sentence 14 contain five, six or seven ideas. Sentence 16 has five ideas; sentence 17, seven; and sentence 18, eight. The rest have more. If this analysis is reasonably close to the facts it would seem to show that the memory span of these children for ideas was at least five and perhaps six or more. This is about the same as, or perhaps a little more than, the average for the digits test, which was the longest of the four Gates subtests.

The intercorrelations of this sentence test with the tests of the other types of memory span, (Table 1) were practically the same as those of the others, with two exceptions. The intercorrelation of the sentence test with the mixed syllables test, .68, is higher than those of the other types of tests and mixed syllables (column 6, Table 1). The intercorrelation with letters, .42, is lower than those of the other tests and letters (column 3, Table 1). These differences may be chance differences due to the small number of cases.

In conclusion, regarding the effect of the kind of test material used, it seems that analysis of intercorrelations and averages gives no grounds for believing that close similarity of content material of the tests produced any appreciable similarity in measurable abilities of so-called memory span.

The order of the length of average memory spans for 18 of the former 25 grade I children when retested in the spring of their grade III year was, with one exception, the same as on the first test. The digit span was the longest, the letter span was second instead of third, the mixed syllables third instead of second, familiar words next, and nonsense syllables shortest. There is substantial agreement between the critical ratios for the Grade III and Grade I test results. The

Table 4a

Averages, S. D.'s and P. E.'s of Averages for 18 of the Same Children Retested with the Gates Subtests after 27 Months

	Average	S. D.	P. E. Average
Digits	6.11	.890	.142
Letters	5.78	.989	.157
Nonsense syllables	4.72	.711	.113
Familiar words	5.11	.890	.142
Mixed syllables	5.56	.848	.135
Total *	20.22	2.844	.452

*Total digits, letters, familiar words, and nonsense syllables tests.

Table 4b

Critical Ratios of the Differences Between the Averages of the Gates Memory Span Subtests for 18 of the Children Tested 27 Months Apart

Differences between	Critical Ratios Spring 1936 Grade III	Dec. 1933 Grade I
Digits and Letters	1.06	2.45
" " Nonsense Syllables	5.15	3.71
" " Familiar Words	3.37	3.92
" " Mixed Syllables	1.90	1.18
Letters and Nonsense Syllables	3.71	1.27
" " Familiar Words	2.15	1.19
" " Mixed Syllables	.72	1.05
Nonsense Syllables and Familiar Words	1.44	.27
" " " Mixed Syllables	3.19	2.18
Familiar Words and Mixed Syllables	1.55	2.19

Table 4c

Differences and Critical Ratios Between the Averages of the Grade I and Grade III Scores made by the 18 Children in the Gates Subtests

	Averages Grade I	Grade III	Differences	S.D. differences	Critical ratios
Digits	5.78	6.11	.33	.303	1.099
Letters	5.06	5.78	.72	.303	2.376
Nonsense syllables	4.72	4.72	.00		0.000
Familiar words	4.78	5.11	.33	.248	1.331
Mixed syllables	5.39	5.56	.17	.318	0.534
Total*	18.83	20.22	1.39	.880	1.580

*Total digits, letters, nonsense syllables, and familiar words tests.

differences between digits and nonsense syllables and between digits and familiar words remained reliable in Grade III. The difference between letters and nonsense syllables, was reliable in Grade III, but not in Grade I. Perhaps the greater familiarity with letters, which the children had acquired by Grade III, made it easier than it had been 27 months before to use the memory image of the letters in holding the letter span in mind. The difference between nonsense syllables and mixed syllables was likewise reliable in Grade III, although it had not been so in Grade I. Familiarity of the symbols could scarcely have been a factor in this instance. Possibly this was an instance of unusual chance result. No reliable differences between Grade I and Grade III abilities is shown by these ratios.

C - CORRELATIONS WITH OTHER MEASURES

Table 5 gives the ninety-two computed correlations of the memory span test with a group of tests, which logically seem to belong together under the general descriptions of "mental tests". The mental measures included parts of the Van Wagenen test, such as information, relationship, and vocabulary tests; all the subtests of the Metropolitan test; the total vocabulary test; performance tests; Binet mental age; and Binet intelligence quotient. The average of all the correlations as derived from the frequency table, was .448; the standard deviation, ±.160. The averages of the correlations of the several memory span tests with the mental tests ranged from .276 to .575.

The averages of the various memory span tests shown were not derived in every instance from correlations of exactly the same mental measures, but that fact does not explain the variations in the sizes of the several averages. Even greater variations appeared among the correlations of particular mental measures with the various memory span tests. For example, the Metropolitan copying test produced correlations with the seven memory span tests ranging from .28 to .72. The Metropolitan Total gave correlations varying from .32 to .68.

It might be supposed that the averages of Table 5 would show tendencies for certain types of memory span material, rather than others, to be more closely related to abilities measured in mental tests. In order of size the averages of the correlations for the different types of test were: sentences, .575; digits, .562; mixed syllables, .470; familiar words, .421; letters, .417; and nonsense syllables, .276. The sentence and digit tests gave the highest averages. Possibly the materials of these tests were more closely related to the materials of the mental abilities tested than were those of the other types of memory span. However, when the individual correlations are considered the variations and exceptions were very

Table 5

Correlations of Memory Span Tests with Mental Measures*

		Gates Reading Diagnosis Tests					
	Van Wagenen Sentences	Digits	Letters	Nonsense Syllables	Fam. Words	Total Col.2-5	Mixed Syllables
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Van Wagenen, Information	.58	--	.32	.15	.45	.28	.55
Van Wagenen, Relations	.64	--	.30	.21	.34	.30	.61
Van Wagenen, Vocabulary	.70	.55	.32	.35	.37	.43	.61
**Metropolitan, Similarities	.39	.63	.46	.26	.45	.51	.26
Metropolitan, Copying	.47	.72	.49	.28	.41	.54	.35
Metropolitan, Vocabulary	.59	--	.39	.32	.52	.34	.66
Metropolitan, Sentences	--	.43	.26	.13	.58	.31	.36
Metropolitan, Numbers	.58	.54	.44	.33	.46	.50	.75
Metropolitan, Information	.66	.50	.47	.42	.41	.50	.43
Metropolitan, Total	.68	.67	.55	.32	.54	.58	.67
Metropolitan, Drawing	--	.55	.37	.25	.62	.41	.39
Vocabulary	.53	--	.29	.24	.05	--	.55
Seguin Form Board, Time	--	.32	.36	.11	.30	--	.07
Mare and Foal, Time	--	.73	.66	.33	.47	.68	.25
Mare and Foal, Errors	--	.54	.41	--	.44	.42	.20
Mental Age	.54	.55	.53	.29	.39	.49	.72
Intelligence Quotient	.54	.57	.46	.43	.36	.53	.56
Averages	.575	.562	.417	.276	.421	.455	.470
Ranges	.39 to .70	.32 to .73	.26 to .66	.11 to .43	.05 to .62	.28 to .68	.07 to .75

*The dashes in the columns of the table mean that the correlations were not computed because the "finder" device indicated that they would not be high.
**Metropolitan Reading Tests for First Grade and Kindergarten.

Frequency Table of Above Correlations, Excluding Totals

.70 to .79	5	.30 to .39	21
.60 " .69	11	.20 " .29	11
.50 " .59	21	.10 " .19	3
.40 " .49	18	0 " .09	2
			92

Ave. = .448
S. D. = ± .16

great. For example, the computed correlations for the sentence and mental tests ranged from .39 to .70; those for the digits test from .32 to .73. The highest individual correlation in Table 5, .75, was for the mixed syllables and Metropolitan numbers tests.

The average of the correlations of the nonsense syllables and the mental tests, .276, was much lower than those for either sentences or digits and the mental tests. Sixteen coefficients were computed and ranged from .11 to .43. The reliability of the nonsense syllables tests as indicated by the average of five intercorrelations, (Table 1, column 4), was a little lower than those for sentence and digits tests. This lower reliability of the test might have reduced the size of the correlations with the mental tests below those of the sentence and digit tests. This explanation loses force, however, when the results of the mixed syllables test are examined. That test had practically the same average for intercorrelations (Table 1, column 6), as did the nonsense syllables. Its correlations with the mental tests, however, averaged nearly as much, .470, as did the sentence and digit tests, .575 and .562, respectively, although the variability was greater, .07 to .75. It seems that many uncontrolled factors operated in the test situations.

In Table 6 are shown the sixty-one computed correlations of the memory span subtests, and eleven correlations of the total memory span test, with a variety of reading tests. The average of the sixty-one correlations of this group, as derived from the frequency table, was .357, the standard deviation ±.151. The averages of the several memory span tests ran from .188 to .546. These averages are a little lower than those for the mental tests. There seemed to be tendencies for the subtests for digits and mixed syllables to correlate higher with reading tests than the other subtests. The averages of the coefficients for each of these tests were about 20-30 points higher than the averages for the letters, familiar words, nonsense syllables and sentences tests. Similarity in the content of the sentence and word memory span subtests and that of the reading tests did not affect the size of the correlations of those tests with reading. In fact, some of the coefficients for the correlations of reading with sentence and word memory span tests were quite small. For instance, the teacher's estimate of pupils' ability to read in May gave a correlation of only .24 with the word memory span test, and .23 with the sentence test, whereas it was .42 with the digits, .38 with the letter tests, and .49 with the mixed syllables test. The May Gates Primary Reading Test, Type 3, Paragraph reading, gave .26 with the

Table 6

Correlations of Memory Span Tests with Reading Tests*

		Gates Diagnosis Reading Tests					
	Van Wagenen Sentences	Digits	Letters	Nonsense Syllables	Familiar Words	Total Col.2-5	Mixed Syllables
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
**Gates Primary Reading Tests Type 1, November	--	--	.11	.01	.31	-.02	.50
Gates Primary Reading Tests Type 1, March	.20	--	.14	-.01	.06	--	--
Gates Primary Reading Tests Type 2, March	--	.50	.55	.24	.45	.48	--
Gates Primary Reading Tests Type 3, March	--	.42	.51	.22	.43	.43	.51
#Hildreth, Matching Words	.43	.63	.33	.31	.37	.46	--
Hildreth, Matching Sentences	--	.46	.18	.19	.39	.30	--
Hildreth, Matching Words and Phrases in Sentences	--	.54	.42	.20	.31	.40	--
Hildreth, Total	.40	.59	.35	.30	.35	.44	--
Teacher's Ranking, November	.46	.49	.41	.21	.32	.40	.65
Teacher's Ranking, May	.23	.42	.38	.18	.24	.34	.49
Gates Primary Reading Tests Type 1, May	.34	.58	.43	.21	.30	.43	.58
Gates Primary Reading Tests Type 2, May	--	--	.38	.15	.12	--	--
Gates Primary Reading Tests Type 3, May	.33	.49	.36	.24	.26	.41	--
Averages	.341	.512	.35	.188	.301	.37	.546

*(See footnote - table 5)
**Gates Primary Reading Tests - Type 1, Word Recognition; Type 2, Sentence Reading; Type 3, Paragraph Reading.
#Hildreth First Grade Reading Analysis Tests

Frequency Table of Above Correlations, Excluding Totals

.60 to .69	2	.20 to .29	10
.50 " .59	9	.10 " .19	7
.40 " .49	14	.00 " .09	2
.30 " .39	16	-.00 "-.09	1
			61

Average = .357 S.D. = ± .151

familiar word memory span test, .33 with sentences, .49 with digits and .36 with letters. The same contrast is shown by the May Gates Primary Reading Tests, Type 1, word recognition, which gave correlations of .30 with familiar words, and .34 with sentence memory span tests, as compared with .58 for the word recognition test and digits and .43 for the word recognition test and letters. The large P. E.'s reduce the statistical significance of the differences between these coefficients.

Table 7 gives the forty-nine computed correlations of the memory span subtests with a variety of letter tests and the seven correlations of the Gates total with letter tests. The average of the forty-nine correlations as derived from the frequency table was .37; the standard deviation ±.167. The averages for the various memory span subtests ran from .228 to .586. The averages of the several memory span tests with various tests of letter abilities show one possible tendency, namely that the familiar word and the nonsense syllables memory span tests correlated lower with letter tests than did the other memory span subtests. For example, the correlations of the familiar words tests with the tests of recognition of phonic combinations, capital letters and small letters, were very low, .14, .18 and .18, respectively. The averages for the Van Wagenen sentences and the Gates digits and mixed syllables tests may be higher than would have been the case if all the correlations had been computed. Columns 1, 2 and 7 in table 7

Table 7

Correlations of Memory Span Tests with Letter Tests*

Letter Tests	Van Wagenen Sentences	Digits	Letters	Nonsense Syllables	Fam. Words	Total Col.2-5	Mixed Syllables
		Gates Reading Diagnosis Tests					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
**Stone, Pt.I	--	--	.46	.19	--	.26	--
#Gates Subtests							
1X, 1-7 Phonic combinations	--	--	.28	.20	.14	.24	--
VIII, 3 Word recognition	--	--	.26	--	.23	--	--
VIII, 2 Word recognition	.50	.52	.42	.44	.39	.50	--
IX, 9 Giving letter sounds	.45	.41	.39	.43	.31	.45	.56
X, 1 Recognition blended words	--	--	.13	.17	.25	--	--
X, 2 Recognition sounded letters	--	--	.09	.20	.03	--	.47
X, 3-4 Giving initial and final sounds	--	.47	.51	.27	.23	.42	.69
XIII,1-2 Writing words	.34	--	.23	.22	.34	.30	.79
XIII, 3 Writing letters	.48	.59	.54	.38	--	.58	.61
1X, 10 Recognition capital letters	--	.40	.52	.19	.18	--	.43
1X, 11 Recognition small letters	--	--	.45	.33	.18	--	.55
Averages	.493	.478	.357	.275	.228	.393	.586

*See footnote in Table 5.
**Stone and Grover Classification Test for Beginners in Reading
Gates Reading Diagnosis Tests

Frequency Table of Above Correlations, Excluding Totals

.70 to .79	1	.20 to .29	10
.60 " .69	2	.10 " .19	7
.50 " .59	9	.00 " .09	2
.40 " .49	12		49
.30 " .39	6		

Average .37 S.D. ± .167

show that many of the correlations of the letter tests with those memory span tests were not computed because the "finder" device indicated that they would be low. The wide total range, .03 to .79, points to the same probability. Had all or most of the missing correlations in those columns been computed the averages might have been considerably lower.

The average of the correlations of the memory span subtest for letters and the twelve letter tests (Column 3) was .357, not as high as four of the other averages. The range was from .09 to .54. The correlation with phonic combinations was .28, with recognition of capital letters .52, and with small letters .45. Similarity of content seemed to have no particular effect on the size of these correlations.

Table 8 shows thirty-four computed correlations of the memory span subtests with a variety of psycho-physical measures and three correlations of the total memory span test with psycho-physical measures. The averages of all computed by the frequency table was .263; the standard deviation ±.233. The several averages for these tests ran from .127 to .375. Only a few of the coefficients in this table were even fairly high. It seems that slight or no relationships existed between the traits classified in this grouping and the memory span abilities measured.

In Table 9 appear the fourteen correlations of the memory span tests with reversals, and the seven correlations with a total error score. The relationships shown are in terms of fewest reversals and errors compared with longest memory spans. Reversals were of the usual variety common to grade I, including letters, words and digits, and "full" and "partial" reversals. They were divided into two groups, those occurring when the stimulus was visual and those occurring when the stimulus was auditory. Nine tests were used in counting the number of visual,

Table 8

Correlations of Memory Span Tests with Psycho-Physical Measures*

	Gates Diagnosis Tests						
	Van Wagenen Sentences	Digits	Letters	Nonsense Syllables	Fam. Words	Total Col.2-5	Mixed Syllables
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
**G. D. XIII, 3 Writing Time	.36	.40	.40	.42	.40	--	.57
Total Vocabulary Time	--	--	.60	.44	.50	--	.42
Perception	.21	.37	.15	.31	.29	.29	-.08
Steadiness	--	--	.54	.05	.62	--	--
Tapping	--	--	.16	.48	.64	--	--
Perseveration	.05	.32	.29	.15	.04	.17	-.10
Chronological Age	.00	.00	.19	-.23	.14	-.08	.30
Averages	.13	.272	.333	.231	.375	.127	.222

* (See footnote - table 5)
** Gates Reading Diagnosis Tests

Frequency Table of Above Correlations, Excluding Totals

.60 to .69	3	.10 to .19	5
.50 " .59	3	.00 " .09	4
.40 " .49	7	-.00 " -.09	2
.30 " .39	5	-.10 " -.19	1
.20 " .29	3	-.20 " -.29	1
			34

Average = .263 S.D. = ± .233

Table 9

Correlations of Memory Span Tests with Reversals and Total Errors

		Gates Diagnosis Tests						
	Van Wagenen Sentences	Digits	Letters	Nonsense Syllables	Fam. Words	Total Col.2-5	Mixed Syllables	Ave.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	
Reversals, Visual	-.04	.30	.22	.18	.25	.26	.10	.18
Reversals, Auditory	.08	.40	-.02	.38	.25	.37	.09	.22
Averages	.02	.35	.10	.28	.25	.315	.145	
Total Errors	.54	.51	.36	.23	.30	.40	.60	.42

Frequency Table of Correlations with Reversals

.40 to .49	1
.30 " .39	2
.20 " .29	3
.10 " .19	2
.00 " .09	2
-.00 "-.09	2
	12

Average = .161
S.D. = ± .155

and five in counting the number of auditory, reversals. The error score was the number of incorrect responses, not including omissions, in six reading tests, fourteen subtests of the Gates Reading Diagnosis Tests, a battery of vocabulary tests, the Mare and Foal test and an original perception test. The correlations of the memory span sub-tests with fewness of errors were moderately high, averaging .42, and ranging from .23 to .60. Those of memory span with reversals were low, averaging .161 with a standard deviation of ±.155.

Most of the reversals counted were made with letters. The correlations of the letter memory span test and reversals averaged .10, indicating no apparent relationship between memory span for letters and ability to write and read letters, numbers, and words without making reversals.

For convenience Table 10 assembles, by types, the average correlations of the various memory span tests with the groupings of mental, reading, letter, psychophysical, reversal and error measures.

Table 10

Averages of Correlations of Memory Span Tests with Groups of Other Measures

		Gates Reading Diagnosis Tests						
	Van Wagenen Sentences	Digits	Letters	Nonsense Syllables	Words	Total Col.2-5	Mixed Syllables	Ave.Col. 1-5 & 7
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mental	.58	.56	.42	.28	.42	.46	.47	.42
Reading	.34	.51	.35	.19	.30	.37	.55	.40
Letters	.49	.48	.36	.28	.23	.39	.59	.43
Psycho-physical	.13	.27	.33	.23	.38	.13	.22	.27
Reversals	.02	.35	.10	.28	.25	.32	.15	.12
Errors	.54	.51	.36	.23	.30	.40	.60	.48

CONCLUSIONS

(1) Averages for grade one children showed that statistically the memory span for digits was longer than those for familiar words, nonsense syllables and mixed syllables. After two and one-half years the digit span remained the longest on retests. The span for letters was almost as long as that for digits for grade one children, and for grade three children was reliably longer than the span for nonsense syllables. Possibly memory images of visual forms aided the children in the digits and letter tests. The span for ideas as expressed in sentence form seemed to be about as long as the digit or letter spans for grade one children.

(2) The differences in size of the averages of the correlations of the various types of memory span tests with other tests do not have great statistical significance, since the probable errors of the correlations from which the averages were obtained were relatively high, as shown in Table 2. However, comparison of averages of these correlations by types of test show possible tendencies for the digit, sentence and the mixed syllables memory span tests to be more closely related to mental, reading, and letter tests and to fewness of errors than were letter, nonsense syllable, and familiar word memory span tests. If such tendencies did exist they were probably due to uncontrolled and complicating factors in the test situations rather than to the nature of the materials of the different tests.

(3) It seems clear, also, that memory span ability is not a unitary ability for young children. Perhaps there is a common core or "general" memory span ability, but, if so, it expresses itself with other and varying abilities which results in much variation in the results.

(4) From the preceding conclusions it would seem that experiences and resulting learnings have considerable effect upon memory span abilities, but relatively much less effect than maturation of mental ability with age.

A STUDY OF CHILDREN'S REACTIONS TO FAILURE AND AN EXPERIMENTAL ATTEMPT TO MODIFY THEM[1]

MARY E. KEISTER AND RUTH UPDEGRAFF[2]

Psychologists and educators believe that it is important for an individual to respond adequately in situations involving failure or great difficulty. After his first attempt meets failure, the individual's subsequent, possibly characteristic reaction is related not only to his emotional adjustment but also to his ability to learn and to profit by experience.

It is natural for a young child to be confronted with many situations which are not readily resolved. Moreover, in his attempts to meet and overcome difficulties as they arise lie the child's opportunities to learn. In general, mental hygienists and educators have considered it desirable for a child to attack a difficult problem with composure, to try out one possibility after another in an attempt to reach a solution. It is usually considered that he is not meeting the situation desirably if he retreats from the problem, if he rationalizes, if he leans heavily on an adult for assistance, if he attacks the problem with such emotional accompaniments as crying, sulking, and tantrums.

Even the most casual observation of young children reveals wide differences in such responses. In the face of a difficult situation, some children make attempts at their own solution, intently and without emotion. There are others, however, who under many circumstances, immediately ask for the help of an adult or another child; some retreat from the scene of action when they discover difficultv; some cry or become angry; some rationalize.

Given, then, a variability from child to child (and in some cases the occurrence of modes of behavior which are undesirable from the standpoint of the future as well as of the present), the problem becomes one of discovering the existence of an undesirable pattern and of modifying that pattern if possible. Such was the problem of this study, the purpose of which may be summarized as follows:

1. To devise tests by means of which one may discover what responses a child of preschool age gives when faced with failure.
2. To select a group of children evidencing undesirable modes of response.
3. To attempt to modify, by special help or individual training, the responses of the children in this group.

Mental hygienists have employed the concept of failure in two ways. They have used it in connection with a situation which is ultimately impossible for the individual to overcome because of his own incapacity; under such circumstances it is important for him to realize this fact and adjust himself to the idea of the impossibility. In the second sense, failure has been thought of as a step in the

[1] A more detailed account of this study may be found in Updegraff, Ruth, Keister, Mary Elizabeth, Heiliger, Louise and others: Studies in Preschool Education I. Univ. Iowa Stud., Stud. in Child Welfare, 1937, 13, No. 4. (In press)

[2] From Iowa Child Welfare Research Station, State University of Iowa, Iowa City, Iowa.

process of solving a problem, as involved in the individual's working his way out of a difficulty. It is with behavior of the latter type that this study is concerned. Failure, as defined here, is the child's lack of immediate success following an attempt to contend with a situation, the situation being one in which he sees some relation to himself as an instrument of his own success or failure.

A preliminary survey of suitable approaches indicated the inapplicability of the observational method, at least in the beginning stages of the study. Not only did it become apparent that failure situations occurred in the nursery school with such infrequency that the time-sampling method was too extensive, but also controls of motivation and of the difficulty of the tasks were lacking. Accordingly, plans were made for presenting failure in experimental situations. The decision was made to confront the child with one situation somewhat in the form of a puzzle, with another which challenged his physical strength, and with a third which offered social obstacles. Among the criteria for setting up the experiments were the following:

1. They must be possible of accomplishment and yet of such difficulty that the child does not succeed immediately.
2. They must provide situations which are natural, in the sense that the difficulties are not obviously or forcibly imposed.
3. The average child should be able to see for himself that he has failed and to see in the situation some relation to himself as an instrument of his success or failure.

As a result of preliminary study, two test situations were believed adequate for use. The first, the puzzle box test, confronted the subject with a small, lidded, colored box, 9 by 7 by 1 1/2 inches. The box being opened, it was found to have a false bottom within 1/4 inch from the top. On this lay ten small, colored figures, of irregular shape, 1/4 inch thick, representing various objects of interest to children, such as a sailboat and an engine. Because of their form they fitted rather closely into the available space. The experimenter then removed the figures and gave the test instructions which invited the child to put the blocks into the box so that the lid could go down again. In spite of the fact that there were several ways in which the blocks could be fitted into the space, the task was quite a difficult one to complete in the fifteen minutes allowed. There was no question of its being an interesting one to children.

The weighted box test consisted of a five-sided box, weighted at the ends and through the middle with from 60 to 90 pounds of iron weights. These weights were adjustable. The box was placed in the middle of a room upside down over a group of attractive toys. When the subject entered, the box was raised slightly, then lowered. Instructions indicated that the toys could be played with if the box could be lifted in order to obtain them. Ten minutes was the time allowed.

The same scheme for recording behavior, a system of controlled observation with time divisions of minutes, was used for both tests. The type of behavior observed is indicated in the tables.

The subjects in this study, 82 children (38 boys and 44 girls) aged three to six years, were enrolled in the preschool laboratories of the Iowa Child Welfare Research Station. The mean intelligence quotient was 122. Because the tests evidenced no statistically significant age differences, marked individual differences being apparent at all ages, the data have not been classified into age

groups. Comparative frequency of various types of responses in the two tests are indicated in Tables 1 and 2. In each test the most frequent response of the group as a whole was "attempts to solve alone" although "interest" ran a close second. That requests for either partial or complete help and rationalizations were more common than disgruntled emotional responses proved to be the case.

Inasmuch as it was the purpose of these tests to differentiate between those subjects giving undesirable or immature responses and those responding more desirably, the extent to which this end was achieved was first to be determined. To describe the process briefly, certain objective criteria were set up in terms of test behavior. Five kinds of behavior occurring for at least a minimum amount

Table 1

Mean Number of Minutes During Which Responses Occurred During Puzzle Box Test (N = 81)*

Behavior	Mean	Standard Deviation
No overt attempt	2.2	3.2
Attempts to solve alone	11.1	4.2
Asks another to solve	1.2	2.3
Asks help	1.5	2.1
Destructive behavior	.1	.5
Rationalizes	1.2	1.8
Interest	10.2	4.7
No emotional manifestations	1.6	2.9
Indifference	.2	1.4
Smiles	.2	.9
Laughs	.1	.2
Sulks	.2	.6
Cries	.3	1.2
Whines	.8	2.0
Yells	.1	.4
Motor manifestations of anger	.04	.3

*Mean length of experimental period: 13.3 minutes.

Table 2

Mean Number of Minutes During Which Responses Occurred During Weighted Box Test (N = 74)*

Behavior	Mean	Standard Deviation
No overt attempt	3.4	2.8
Attempts to solve alone	5.7	2.7
Asks another to solve	.4	3.6
Asks help	1.1	1.9
Rationalizes	1.0	1.5
Interest	5.7	3.2
No emotional manifestations	2.1	2.6
Indifference	.1	.9
Smiles	.3	.8
Laughs	.1	.4
Sulks	.2	1.0
Cries	.3	.9
Whines	.7	1.7

*Mean length of experimental period: 9.1 minutes.

of time were listed and definitely stated quantitatively. If a child's behavior fell into two or more of these classifications on either or both tests, he was judged to have given an immature response. In brief, these five types were as

follows: (1) giving up attempts to solve the puzzle box in less than five minutes or to solve the weighted box in less than two minutes, (2) requesting help during more than one-half the total time of the test, (3) manifesting destructive behavior, (4) making more than two rationalizations, (5) evidencing exaggerated emotional responses.

Analysis of the test records showed a total of fifteen children (18 per cent) who fell into the immature group.

The diagnostic value of the tests is illustrated by contrasting frequencies of behavior as shown in Tables 3 and 4, in which it is apparent that real differences do exist between the groups as classified by this means.

The next step in the study was the training program. In this, twelve out of the fifteen children participated.

Table 3

Mean and Standard Deviation of Responses in Minutes for Two Groups of Subjects on Puzzle Box Test

Behavior	Group Showing Undesirable Or Immature Response (N = 15)		Remainder of Total Group (N = 54)	
	Mean	Standard Deviation	Mean	Standard Deviation
No overt attempt	6.0	3.7	1.6	2.3
Attempts to solve alone	8.5	4.2	13.0	3.0
Asks another to solve	3.6	3.4	.8	1.6
Asks help	2.5	2.4	1.5	2.1
Destructive behavior	.6	1.1		
Rationalizes	2.8	2.5	1.0	1.4
Interest	6.0	3.8	12.4	3.9
No emotional manifestations	2.5	2.4	1.7	3.2
Sulks	.8	1.3		
Cries	1.7	2.4		
Whines	2.6	2.9	.5	1.5
Yells	.3	.8		
Motor manifestations of anger	.2	.5		

Table 4

Mean and Standard Deviation of Responses in Minutes for Two Groups of Subjects on Weighted Box Test

Behavior	Group Showing Undesirable Or Immature Response (N = 15)		Remainder of Total Group (N = 50)	
	Mean	Standard Deviation	Mean	Standard Deviation
No overt attempt	5.2	2.9	3.4	2.4
Attempts to solve alone	4.2	2.9	6.5	2.3
Asks another to solve	.7	1.6	.3	.9
Asks help	2.2	2.3	.9	1.7
Rationalizes	1.7	1.2	.8	1.5
Interest	3.5	3.0	6.8	2.9
Sulks	1.0	2.0		
Cries	1.0	1.6	.1	.4
Whines	2.3	2.3	.3	1.2

The basic philosophy underlying the training assumed that children can learn to meet difficulty in a controlled manner and acceptably if they know from experience what type of behavior is most likely to bring success or satisfaction. It was the aim of the training program to raise the responses of the immature

group nearer to the level of desirability. Specifically, in the training an attempt was made to teach the child to persist longer in the face of difficult tasks (which were, however, not impossible ones), to teach him to depend less upon an adult for help, and to attack a problem and see it through with some composure.

The method of training consisted in introducing the child to a series of problems which grew progressively more difficult as the program of training proceeded. The problem situations reflected the following criteria:

1. The tasks should be graded in difficulty so that the child experiences success in the earlier ones and gradually works up to problems which are difficult for him.
2. The later tasks must be of such difficulty that the child does not succeed immediately but is forced to persevere, to continue to try if he is to attain success.
3. The child must be able to see his progress and previous successes.

In describing the two training situations briefly[1] it may be said that they were similar in type but differed in the specific materials used. For the first, four picture-puzzle books were prepared, each one in the series more difficult than the one preceding and each one of graduated difficulty from beginning to end. For these, interesting, colorful and appropriate story books were cut up. The pictures were mounted on 4-ply wood, varnished, cut into puzzles, and the book was rebound on loose rings. The experimenter read the story to the child. As she reached a part illustrated by one of the pictures, she stopped for him to put the puzzle together before continuing the story. After the first picture was completed she covered it with cellophane, so that both she and the child could refer to it later, and resumed the story until the next picture. Each book contained four to six pictures.

In the second situation a "block boy" was built. Copied from a drawn pattern hung on the wall, he was to be made of colored blocks placed upon each other so that having attained first feet, then legs, then trunk and arms, then head, he stood approximately three feet high, a somewhat precarious figure and a frequently exasperatingly unsteady one. Usually several attempts were necessary in order to complete him. After a successful production his builder had the task of devising a hat from a wide variety of materials provided.

The entire program of training was handled by one person. Training periods varied in length from eight to thirty-three minutes, depending largely upon the difficulty of the tasks and the child's behavior. To subject the twelve children to all of the training took approximately six weeks.

Behavior during the training program underwent a gradual improvement as is shown by both objective and subjective estimate. In order to study post-training behavior objectively, two approaches were utilized; first, retests by means of a similar but not identical puzzle box were given the trained subjects (see Figure 1); second, also retested were an equal number of children, not in the trained group, who during the initial tests had shown some undesirable behavior (see Figure 2).

[1] Detailed descriptions of all the materials used in this study may be obtained from the Iowa Child Welfare Research Station, Iowa City, Iowa.

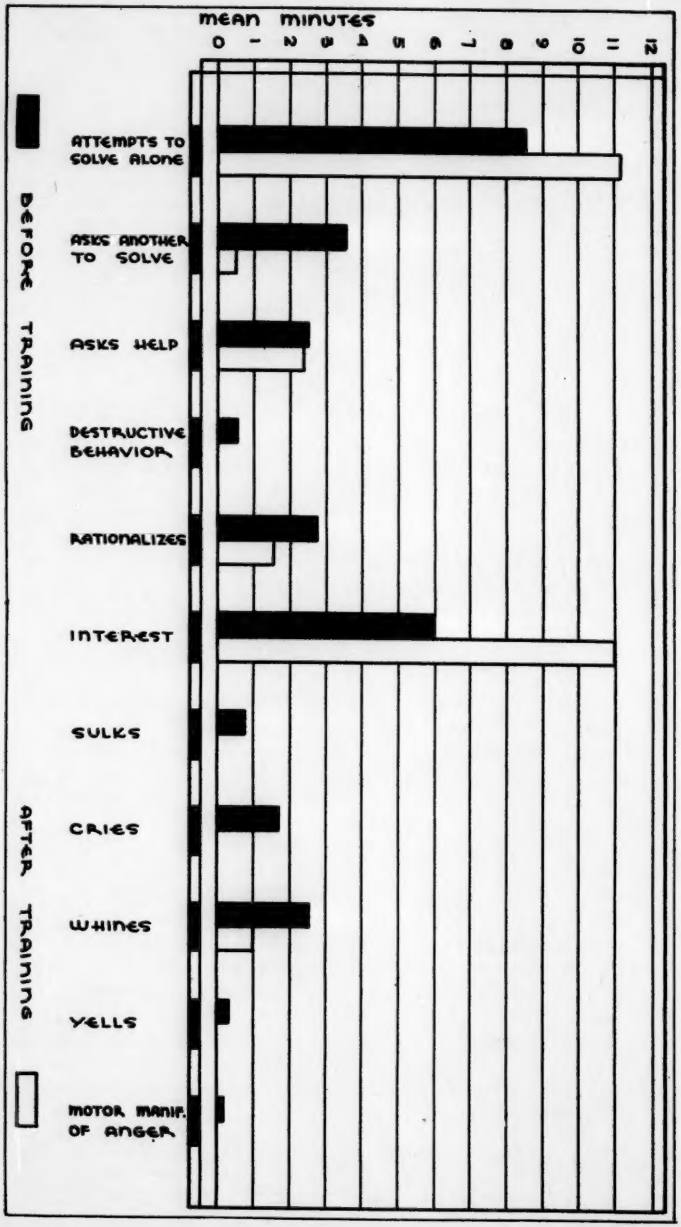


Figure 1. Responses of Trained Group on Puzzle Box Test Before and After Training.

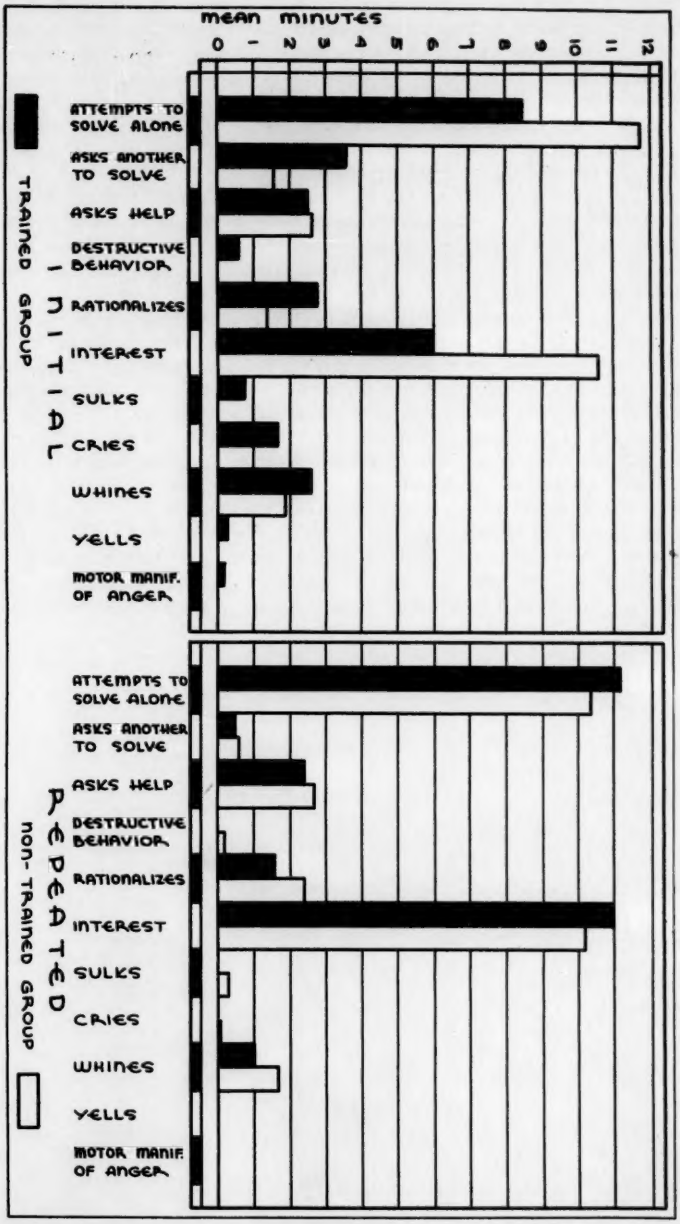


Figure 2. Response of Trained and Untrained Groups on Puzzle Box Test.

It is evident from a study of Figure 1 that the behavior of the children after training was remarkably different from their behavior prior to training. Differences in the three items attempts to solve alone, interest, and cries are statistically significant. Excepting in the case of the item asks help, the remaining differences closely approximate significance. The differences were in form of the responses given in the retest and indicate that a reasonable improvement was effected in the trained group. The exaggerated emotional responses of sulking and crying dropped out entirely in this group.

Figure 2 concerns responses of the trained and the compared nontrained group before and after training. The two groups differed in the responses no overt attempt, attempts to solve alone, interest, sulks, and cries. All of the differences were in favor of the trained subjects in spite of the fact that previous to training the difference lay in the opposite direction.

The results of this study, hopeful as they are, must be interpreted in the light of the specific conditions. The entire program was carried out by the experimenter, who also gave the retests. Further study, at present underway, must determine the extent to which the more desirable behavior occurs in other situations and with other persons. There is evidence that behavior of children in difficulties has been similar in two test situations; it would be valuable to make observations in other situations and under circumstances of a more social nature. Probably the most important contribution of the present study is its indication of the marked effect of this training program. After training, children tried longer, manifested more interest in solving problems themselves, and completely eliminated emotional behavior. Evidently this improvement was not a function of age or other training. Of particular interest to teachers and psychologists may be the fact that the program of training was neither arduous nor time-consuming.

THE OBJECTIVITY OF ANTHROPOMETRIC MEASUREMENTS TAKEN ON EIGHT- AND NINE-YEAR-OLD WHITE MALES

EVERETT L. MARSHALL[1]

PURPOSE

This study has been developed in such a manner as to constitute a companion paper to one recently completed (2). The earlier paper dealt with the reliability of anthropometric measurements - reliability being determined by the differences between initial and repeated observations made by a single anthropometrist. The present study pertains to the objectivity of anthropometric measurements - objectivity being considered as a function of like differences when the initial measurements are made by one anthropometrist and repeated observations are taken by a second anthropometrist.

It follows, then, that the purpose of this study is in part investigative and in part comparative. On the one hand it attempts to investigate the objectivity of a series of physical measurements made on boys eight and nine years of age, and on the other hand it affords a comparison of the resulting objectivity findings with earlier findings for reliability (2). Specific major aims are: (1) To determine to what extent, if at all, one investigator tends to secure smaller or larger measurements than the other for any given dimension. In question form this aim becomes: Are there systematic differences in the measurement technique of the two anthropometrists? (2) To obtain for each of the fifteen anthropometric dimensions objectivity tables. These tables are to be derived from measurement observations taken under relatively optimum conditions. (3) To relate each objectivity table to the rate of growth for the dimension from which it is derived and to compare the resulting "frequency of consecutive measurement" estimates with those obtained for reliability (2).

MEASUREMENTS[2]

The fifteen anthropometric dimensions selected for study are itemized below. Each item is complemented by a descriptive statement of the measurement technique employed in determining the dimension.

Stature: The instruments used were the Baldwin Paper Measuring Scale and Square. The subject stood erect with heels almost touching each other. Heels, buttock, upper part of back, and rear of head were against the wall to which the scale was attached. The arms were permitted to hang at the sides of the body in a natural position, the heels were in firm contact with the floor, and the head was held on the Frankfort Horizontal. One face of the square was so placed against the scale that the other face was horizontal with the floor. The anthropometrist then brought down the square with sufficient force to crush the subject's hair and made the reading.

Sitting Height: The instruments were the same as for stature. The subject sat on a horizontal walnut bench 30 centimeters in height. His knees were flexed and spread apart, his ankles crossed, and his hands rested on his thighs. The posterior aspect of the trunk made contact with the scale both at the sacral region and at the thoracic

[1] From Illinois State Normal University, Normal, Illinois.

[2] The measurement technique described here is supplied verbatim by Dr. Howard V. Meredith. (2)

region. The square was brought down firmly on the vertex (the highest point of the head when held on the Frankfort Horizontal) and the measurement recorded as the distance from this point to the surface of the bench.

Bi-acromial Diameter: Standardized large sliding calipers having broad flat branches were used in taking this measurement. The subject assumed the erect position with his arms hanging at his sides. In an attempt to standardize this diameter as much as possible, the anthropometrist stood behind the subject and drew back the subject's shoulders until the bodies of the two scapulae lay approximately in one plane. The measurement was determined as the distance between the most lateral points of the acromial eminences.

Bi-iliac Diameter: This measurement was taken as the straight distance between the most lateral points of the crests of the ilia. The instrument used was the same as for the previous dimension. The observer stood in front of the subject, brought the face of each branch of the calipers squarely in contact with the landmark, and applied the maximum pressure that could be exerted without pain to the subject. In the event that the subject appeared to turn his trunk and hips at the time the pressure was applied, the measurement was checked.

Bi-trochanteric Diameter: The subject stood erect with legs together and feet parallel. The most external prominences of the greater trochanters were the terminal measurement points. Pressure was applied to the large sliding calipers until considerable resistance of the bones was felt.

Bi-condylar Diameter of Humerus: This measurement was taken on the upper left extremity by means of the large sliding calipers. The left arm of the subject was raised forward to approximately the level of the shoulder and the forearm flexed upward at right angles to the arm. The branches of the calipers were then applied against the epicondyles of the humerus in such a manner as to bisect the angle of the elbow and lie in the plane of the arm and forearm. Heavy pressure was used, care being taken that the forearm did not move out of line with the branches of the calipers and tip in toward the midline of the body.

Bi-condylar Diameter of Femur: Here the large sliding calipers were used and the maximum straight distance obtained between the condyles of the left femur. The knee of the subject was flexed sufficiently to relax and largely remove the musculature at the lateral aspects of the condyles. The branches of the calipers bisected the thigh-leg angle as they were brought in contact with the bony prominences. Considerable pressure was used. The anthropometrist stood in front of the subject while taking this measurement.

Circumference of Thorax: A steel millimeter tape was used to determine this measurement. The subject stood in a natural manner with head erect and with arms relaxed and held slightly away from the sides of the body in order to permit the passing of the tape around the thorax. The observer stood in front of the subject and, using the xiphoid cartilage of the sternum as the anterior landmark, placed the tape around the thorax at right angles to the spinal column. Posteriorly, the tape always rested below the inferior angles of the scapulae. The tension applied was only sufficient to insure complete contact with the skin. Record was made of the median value during the normal respiration.

Circumference of Arm: The instrument employed in securing this and the three succeeding limb girths was the steel millimeter tape. As with thoracic circumference, the tape was applied to make contact all around and yet avoid compression of the tissues. All observations were taken on the left extremities of the body.

Arm circumference was measured near the middle of the humerus, at the level of the greatest girth over the biceps muscle and below the insertion of the deltoid muscle. The plane of the tape was at right angles to the line of the humerus. During measurement the subject assumed the erect position with the upper extremities hanging near the sides of the body (slightly abducted) in a relaxed condition.

Circumference of Forearm: The position of the limb was the same as for the previous measurement, and identical technique was used except that the observation was taken at the level of the greatest girth below the elbow joint and in the region of the radiale. Care was taken to see that the musculature of the forearm and hand was relaxed.

Circumference of Thigh: The subject stood with his feet spread about 9 inches apart and his weight equally distributed on both lower extremities. The tape was passed around the thigh at right angles to its long axis and the measurement made at a level just below the gluteal sulcus.

Circumference of Leg: The subject maintained his position as for the previous measurement and the maximum girth of the calf at right angles to its long axis was determined.

Thickness of Skin and Subcutaneous Tissue at Thorax Back: This measurement, together with the two measurements which remain to be described, was taken with so-called "fat" calipers devised by the American Child Health Association. In all three instances the flat, blunt-nosed branches of the calipers were held parallel to the long axis of the body or extremity of the subject.

In taking the measurement at the rear of the thorax, the anthropometrist placed the thumb and first finger of his left hand about 40 millimeters apart over the region below and slightly lateral to the inferior angle of the left scapula and in the transverse plane

of the xiphoid cartilage. He then moved these digits directly toward each other, taking care that they did not tend away from the thorax. The instrument was next applied to the tissue held between his thumb and finger and the measurement read off. The objective was to measure the thickness of a complete double layer of skin and subcutaneous tissue without including any muscle tissue.

Thickness of Skin and Subcutaneous Tissue at Arm Back: This measurement was taken over the triceps muscle at approximately the mid-point of the shaft of the left humerus with the arm hanging in a relaxed condition. The technique was the same as for the previous measurement.

Thickness of Skin and Subcutaneous Tissue above Iliac Crest: The calipers were applied immediately superior to the crest of the left ilium in a line vertical with the left axilla. The size of the bite taken between the digits of the anthropometrist varied with the amount of subcutaneous tissue of the individual subject.

All measurements were made on the nude subject and recorded to the nearest millimeter.

SUBJECTS

The subjects for this study were twenty-five Iowa City boys enrolled in the third and fourth grades of the University of Iowa elementary school during the year 1935-1936. Each boy was scheduled to serve as a subject for the study six consecutive times, once every four weeks beginning November 13 or 14 and ending April 1 or 2. Occasional absences, however, reduced the total number of examinations to 141. The youngest boy in the group was seven years, eight months at the time the first examination was made, and the oldest boy was ten years, two months by the time of the final examination.

The principal reason for accumulating the data on twenty-five individuals rather than on 141 different individuals lay in the fact that this smaller group was already under observation in connection with a seriatim growth study.

> "In addition to being conveniently available, however, the sample was found to represent considerable dispersion in body size and build. The eight-year-old boys ranged between 128.0 cm. and 143.9 cm. for stature, between 57.2 cm. and 66.6 cm. for thoracic circumference, and between 33.4 cm. and 43.5 cm. for girth of thigh. The shortest boy had the largest thigh circumference and the tallest boy the second smallest girth of thorax. Similar deviation characterized the nine-year-olds. Stature ranged from 132.6 cm. to 152.7 cm., bi-iliac diameter from 19.9 cm. to 24.1 cm., and arm girth from 17.4 cm. to 23.1 cm. The tallest boy had the largest bi-iliac diameter while the shortest had the largest circumference of arm." (2)

EXAMINATION PROCEDURE

The following procedure was employed at each examination:

1. The subject removed all clothing.
2. The first anthropometrist made observations for each of the fifteen measurements being investigated and serially called off each observed value to a recorder.
3. As the first anthropometrist left the room, the second one entered and made observations for the same fifteen physical dimensions.

Two investigators, Dr. Howard V. Meredith and the writer, made all the measurements employed in this study. The recording was done by Dr. Newell C. Kephart.

Extreme care was exercised at all times to measure as accurately as possible and to avoid errors in recording.

CONSISTENT DIFFERENCES

The data consist of 141 pairs of observations for each of fifteen anthropometric dimensions, each pair of which is composed of two successive measurements made by two anthropometrists. To check for any tendency on the part of either investigator consistently to secure smaller or larger measurements for any given dimension than those secured by the other investigator, the following procedure was employed. Considering the measurement secured by one investigator (the writer) throughout as the starting point, the difference between the measurement secured by the first investigator and that secured by the second was marked positive if the latter measurement was larger and negative if it was smaller. Next the mean of the positive differences and the mean of the negative differences were found for each dimension. The difference between these means was then computed; the results are shown in the final column of Table 1. In ten of the fifteen measurements, the mean difference in a positive or negative direction was less than 1 millimeter. Three of the mean directional differences were between 1 and 2 millimeters and two were exactly 2 millimeters. For every dimension having a positive "mean difference" (see Table 1) the positive differences were found to exceed, in number, the negative differences, and for every dimension having a negative "mean difference," there were a greater number of negative than positive differences. This indicates that the mean directional differences are representative and are not the function of a few extreme differences. The second investigator (the writer) secured measurements for height and sitting height which were on the average 1.5 millimeters larger than those secured by the first investigator. For dimensions of thoracic circumference, thigh girth, and arm girth the opposite was true - on the average the measurements secured for the first two dimensions by the second investigator were 2 millimeters smaller and for arm girth 1 millimeter smaller. From Table 1

Table 1

Mean Differences Between the Measurements Secured by the Two Investigators Who Made 141 Pairs of Observations for Each of Fifteen Dimensions

Dimension	Positive Differences	Negative Differences	Zero Differences	Mean Differences (mm.)
Stature	83	45	13	+1.4
Sitting height	85	45	11	+1.7
Thoracic circumference	54	81	6	-2.0
Bi-acromial diameter	67	69	5	- .5
Bi-iliac diameter	43	58	40	- .2
Bi-trochanteric diameter	50	65	26	- .7
Bi-condylar diameter of humerus	33	66	42	- .3
Bi-condylar diameter of femur	33	54	54	- .3
Girth of arm	41	80	20	-1.1
Forearm girth	53	52	26	+ .1
Thigh girth	46	89	6	-2.0
Girth of leg	65	50	26	+ .3
Thickness of skin and subcutaneous tissue:				
At arm back	23	63	55	- .5
At thorax back	9	47	85	- .3
Above iliac crest	22	73	46	- .7

it appears that the directional differences are so small that for practical purposes they may be considered negligible.

OBJECTIVITY FINDINGS

The objectivity constants shown in Table 2 were found by computing the differences, irrespective of sign, between the two component measurements comprising each of the 141 pairs of observations. The median, ninetieth percentile, and maximum difference for each of fifteen dimensions were derived from frequency distributions of these differences. Some of the results shown in this table are:

1. The thickness of skin and subcutaneous tissue on the back of the thorax was measured with the least absolute error. The greatest absolute difference was encountered in measurements for stature, sitting height, and bi-acromial diameter.
2. Measurements with objectivity approximating that of thickness of skin and subcutaneous tissue on the back of the thorax were bi-iliac diameter, bi-condylar diameter of humerus, bi-condylar diameter of femur, and the other two measurements of skin and subcutaneous tissue.
3. Bi-iliac diameter was measured more accurately than bi-trochanteric diameter.
4. The measurements of leg girth and forearm girth (soft tissue dimensions) were secured with less absolute error than bi-trochanteric diameter (a bony dimension).
5. Forearm girth and leg girth were measured with more consistency than thigh girth and arm girth.

Additional findings were secured (1) by calculating the coefficient of objectivity by the Pearson product-moment method of correlation, and (2) by expressing the median and maximum values in Table 2 in relation to the actual size of the dimension from which each was derived. Means for the various dimensions for nine-year-old boys were secured from studies by Meredith (1) and by Meredith and Boynton (3).

From Table 3 it is noted:

1. The objectivity medians in Table 2, when referred to the mean magnitude of their respective dimensions, equal one-half of 1 per cent or less for stature, sitting height, bi-iliac diameter, thoracic circumference, and leg girth; less than 1 per cent for bi-trochanteric diameter, forearm girth, bi-condylar diameter of the femur, and arm girth; less than 2 per cent for thigh girth, bi-acromial diameter, and bi-condylar diameter of the humerus; and from 4 to 9 per cent for the three measurements of skin and subcutaneous tissue.
2. The maximum measurement differences given in Table 2 range from approximately 1 per cent of the mean size of their corresponding dimensions to 60 per cent. Stature has the lowest percentage; the other dimensions, excepting those of skin and subcutaneous tissue whose percentages are 24 or greater, have percentages ranging from 1 to 9.
3. The coefficients of objectivity range from .80 to 1.00, stature having the highest and bi-acromial diameter having the lowest.

OBJECTIVITY FINDINGS IN RELATION TO RATE OF GROWTH

The relationship between objectivity findings and rate of growth was determined by proceeding on the assumption used by Meredith (2) in relating findings for

Table 2

Objectivity Constants for Fifteen Anthropometric Dimensions: Each Series of Constants Was Derived from Values Representing the Difference Between Two Measurements Taken Successively by Two Anthropometrists on the Same Subject

Difference Interval (mm.)	Stature		Sitting Height		Thoracic Circumference		Bi-acromial Diameter	
	Number	Per Cent	Number	Per Cent	Number	Per Cent	Number	Per Cent
0 to 1	38	27.0	33	23.4	35	24.8	40	28.5
2 to 5	58	41.1	73	51.8	60	42.6	57	40.4
6 to 10	36	25.5	22	15.6	31	22.0	24	17.0
11 to 15	7	5.0	10	7.1	10	7.1	15	10.6
16 to 25	2	1.4	3	2.1	5	3.5	5	3.5
	141	100.0	141	100.0	141	100.0	141	100.0
Median	3.6		3.1		2.8		3.0	
90th Percentile	9.2		9.5		11.0		12.5	
Maximum	19.		20.		24.		24.	

Difference Interval (mm.)	Arm Girth		Forearm Girth		Thigh Girth		Leg Girth	
	Number	Per Cent	Number	Per Cent	Number	Per Cent	Number	Per Cent
0 to 1	59	41.9	80	56.7	26	18.5	74	52.5
2 to 5	70	49.6	54	38.3	66	46.8	64	45.4
6 to 10	11	7.8	7	5.0	37	26.2	3	2.1
11 to 15	1	.7			12	8.5		
	141	100.0	141	100.0	141	100.0	141	100.0
Median	1.9		1.3		4.0		1.4	
90th Percentile	5.3		4.1		9.5		4.2	
Maximum	15.		9.		17.		8.	

Difference Interval (mm.)	Bi-trochanteric Diameter		Bi-iliac Diameter		Breadth of Elbow		Breadth of Knee	
	Number	Per Cent	Number	Per Cent	Number	Per Cent	Number	Per Cent
0	26	18.4	40	28.3	42	29.7	54	38.3
1	44	31.2	72	51.1	72	51.1	68	48.2
2	27	19.1	19	13.5	19	13.5	14	9.9
3	18	12.8	8	5.7	8	5.7	3	2.1
4 and 5	16	11.3	2	1.4			2	1.4
6 to 10	10	7.1						
	141	99.9	141	100.0	141	100.0	141	99.9
Median	1.5		.9		.9		.8	
90th Percentile	5.1		2.3		2.2		1.8	
Maximum	10.		4.		3.		5.	

Thickness of Skin and Subcutaneous Tissue:

Difference Interval* (mm.)	At Arm Back		At Thorax Back		Above Iliac Crest	
	Number	Per Cent	Number	Per Cent	Number	Per Cent
0	55	39.0	85	60.3	46	32.6
1	59	41.8	50	35.5	56	39.7
2	22	15.6	6	4.2	20	14.2
3	5	3.5			11	7.8
4 to 6					8	5.7
	141	99.9	141	100.0	141	100.0
Median	.8		.4		.9	
90th Percentile	2.1		1.3		3.0	
Maximum	3.		2.		6.	

*This series of differences is for the actual measurement of a double layer of tissue. One-half the values given afford an estimate of the errors of measurement for a single layer.

Table 3

Objectivity Findings Supplementary to Those Given in Table 2

Dimension	Median from Table 2 x 100 / Mean Size of Dimension	Maximum from Table 2 x 100 / Mean Size of Dimension	Objectivity Coefficient
Stature	.26	1.39	1.00
Bi-iliac diameter	.42	1.87	.98
Sitting height	.43	2.76	.97
Thoracic circumference	.44	3.73	.97
Girth of leg	.52	2.98	.97
Bi-trochanteric diameter	.63	4.22	.96
Forearm girth	.67	4.67	.95
Bi-condylar diameter of femur	.95	5.97	.93
Girth of arm	.99	7.78	.97
Thigh girth	1.01	4.31	.97
Bi-acromial diameter	1.03	8.22	.80
Bi-condylar diameter of humerus	1.63	5.42	.87
Thickness of skin and subcutaneous tissue:			
At thorax back	4.88	24.39	.93
At arm back	6.25	23.44	.93
Above iliac crest	9.00	60.00	.84

reliability to the rate of growth, namely, that seriatim observations are profitably made only when the interval between them is large enough to allow a mean increment of growth which equals or exceeds 90 per cent of the reliability differences. The reliability differences here become objectivity differences, but otherwise the procedure is similar.

Table 4 shows the findings on relationship for eight- and nine-year-old boys. Given successively in the first three columns are (1) the series of dimensions under consideration, (2) the annual rate of mean growth in each dimension, and (3) the ninetieth percentiles of objectivity distributions for each dimension. The time frequency at which it appears practical to make seriatim observations for the purpose of studying the growth of the individual appears in column four.

Table 4

Relationship Between Anthropometric Objectivity Constants and Growth Increments for Eight- and Nine-Year-Old Boys

Dimension	Mean Annual Increment* (mm.)	Objectivity Criterion (mm.)	Estimated Maximum Frequency of Measurements
Stature	54	9.2	Bimonthly (?)
Bi-iliac diameter	8	2.3	Quarterly (?)
Girth of leg	10	4.2	Semiannually
Bi-trochanteric diameter	10	5.1	Semiannually
Sitting height	21	9.5	Semiannually
Thoracic circumference	22	11.0	Semiannually
Girth of thigh	16	9.5	Semiannually (?)
Forearm girth	6	4.1	Annually
Arm girth	7	5.3	Annually
Bi-condylar diameter of femur	2	1.8	Annually
Bi-acromial diameter	10	12.5	Annually
Bi-condylar diameter of humerus	1	2.2	Biannually

*This column of values was taken from Meredith (2)

SUMMARY AND CONCLUSIONS

Objectivity findings - objectivity being considered as a function of the differences when initial observations are made by one anthropometrist and repeated measurements are taken by a second anthropometrist - are presented for each of fifteen anthropometric dimensions taken on eight- and nine-year-old boys.

Preliminary computation of the mean directional differences yield results which indicate that systematic differences in measurement technique of the two anthropometrists for each of the fifteen dimensions are of negligible importance.

Objectivity tables derived from pairs of measurement observations taken under unusually favorable conditions are presented. Some of the findings are: the most objective dimension is the thickness of skin and subcutaneous tissue on the back of the thorax; the least objective measurements are stature, sitting height, and bi-acromial diameter.

The objectivity constants, chosen to represent given body dimensions, are referred to the annual rate of growth for the dimension from which they were derived. From this procedure, estimates of the practical time interval to allow between seriatim measurements of various dimensions are made. Some dimensions are found to be taken advantageously quarterly or semiannually while for others it appears that no significant contributions to individual growth trends are made by taking measurements oftener than annually or biannually. By comparing results with those for reliability (2), it appears that changing anthropometrists from one observation period to another increased the size of the time interval for four of the dimensions, but for the remaining dimensions the interval between observations remained the same regardless of the change in anthropometrists. From the results of these two studies it appears obvious that not all measurements should be taken at each examination period of a seriatim measurement program.

Finally, objectivity for the fifteen measurement dimensions studied, for boys eight and nine years of age, compares favorably with reliability. The findings also lend additional evidence to substantiate the conclusions made by Meredith (2) that "efficient and economical research on the physical growth of the individual lies (1) in a differential approach to anthropometric dimensions and (2) in employment of unusually rigorous measurement technique."

REFERENCES

(1) Meredith, Howard V.: The rhythm of physical growth: A study of eighteen anthropometric measurements on Iowa City white males ranging in age between birth and eighteen years. Univ. Iowa Stud., Stud. in Child Welfare, 1935, 11, No. 3, pp. 128.

(2) Meredith, Howard V.: The reliability of anthropometric measurements taken on eight- and nine-year-old white males. Child Develop., 1936, 7, No. 4, 262-272.

(3) Meredith, Howard V., and Boynton, Bernice: The transverse growth of the extremities: An analysis of girth measurements for arm, forearm, thigh, and leg taken on Iowa City white children. Accepted for publication in Human Biology.

PERSONALITY CHARACTERISTICS OF JUVENILE DELINQUENTS
II. RELIABILITY OF DIFFERENTIATING TRAITS

MERVIN A. DUREA [1]

INTRODUCTION

The first section (3) of this investigation dealt with a method whereby items were selected from each of four tests of the Pressey Interest-Attitude Tests which best differentiate delinquents from non-delinquents. As a basis for selecting items three criteria were established. The operation of these was illustrated and the consistency of the method demonstrated. By employing the procedure described there emerged from each test a list of items which appeared to be maximally effective in distinguishing between the control and experimental groups under consideration. Seven differentiating items from Test I (things considered wrong) were found; five from Test II (worries, fears, anxieties); six from Test III (likes and interests); and eight from Test IV (kinds of people liked or admired).

While the consistency of the differentiating items seems to be assured by the analysis performed in the first section, it is the object of the second section to examine further their reliability and diagnostic significance.

PROBLEM AND METHOD

To make clear the procedure followed differential items from each test are listed below:

Test I (things considered wrong). carrying a revolver (+), gang (+), being conceited (-), being a snob (-), playing cards (+), bribery (-), prison (+).

Test II (anxieties, fears, worries). jail (+), family (+), death (+), dying (+), sins (+).

Test III (likes and interests). church (+), circus (+), movie star (+), tap dancing (+), joyriding (+), candy (+).

Test IV (kinds of people liked or admired). handsome (+), husky (+), quick (+), well-dressed (+), wealthy (+), good-looking (+), cooperative (-), rich (+).

It will be noted that after each item a plus or minus sign is shown. In general these signs indicate the direction of original responses. That is, when delinquents responded proportionately a larger number of times than non-delinquents to a given item it was denoted with a plus sign. The converse received a minus designation. Only in Tests I and IV do items occur with minus denotations. Operation of plus and minus signs will be clarified subsequently.

[1] From Ohio State University.

CHILD DEVELOPMENT, Vol. 8, No. 3 (September, 1937)

For purposes of determining the reliability of the differential items, the following course has been adopted:

(a) Interest-Attitude Tests constituting each life age grouping of delinquents were re-scored in terms exclusively of the differential items from each test. Subjects were the same groups, 14, 15, 16, and 17-year, that were used in the statistical analysis of section one, similar life age groupings being retained in the present section.

(b) Each life age group of delinquent boys was paired with a control group of non-delinquent boys of the same life age. Responses of the control groups on each of the Interest-Attitude Tests were likewise differentially scored.

(c) Besides the original groups of subjects the technique of differential scoring was applied to an entirely new sample of 115 delinquent boys of mixed life age, the median being 16.6 years.

(d) Although the analysis in section one which resulted in the selection of differential items was confined to life age groupings 14 to 17 years inclusive, it was considered of interest to ascertain the significance of the items for a younger life age group. Hence, a control and experimental group of 13-year subjects have been included in the present analysis.

In each of the four Interest-Attitude Tests the subject is directed to single-cross (X) or double-cross (XX) items depending on the extent and intensity of reaction to them. Thus in deriving differential scores not only were plus and minus denotations employed but single and double crossings as well. An example will make the scoring device clear:
Case W.H.C., a 16-year old delinquent boy. On Test I the subject responded as follows to the differential items: XX carrying a revolver (+), XX bribery (-), X being conceited (-), XX gang (+), XX playing cards (+), X being a snob (-). No response was made to the item, prison. Summing the responses, double crossed items counting two, single crossed one, to positively denoted items +6 is obtained. Making the same computations for items negatively designated -4 is obtained. Adding algebraically the differential weighted score for Test I is 2. On Test II responses are: XX death (+), XX jail (+), X sins (+), XX dying (+). No response is made to the item, family. Since all differential items in this test are positive the differential score is the sum of single and double crossings, this being 7. On Test III the subject responded to only three of the six differential items: X joyriding (+), X circus (+), X candy (+). No response was made to church, movie star, and tap dancing. All items in Test III are positive. Since three items received one cross each the differential score for the test is 3. On Test IV the subject single-crossed five of the eight differential items: husky, quick, wealthy, well-dressed, and rich. Items, cooperative, good-looking, and handsome were ignored. Since all items to which responses were made are positive the differential score on Test IV is 5. Incidentally it should be noted that Test IV contains one minus item, cooperative, which if indicated by the subject necessitates an algebraic sum. To obtain a differential weighted score for the entire instrument the scores of the four tests were totalled. In the present case the total differential score is the sum of 2, 7, 3, and 5, or 17. If any single test

received a negative score, the total differential score was computed algebraically.

The highest possible differential weighted score on the Interest-Attitude Tests is 44 or the sum of all double-crossed positive items. The lowest possible score is -8 or the sum of four double crossed negative items. Differential scores of the control and experimental groups compared, therefore, distribute between the extremes indicated. The essential problem undertaken is one of determining to what extent the differential scores reliably discriminate delinquents from non-delinquents.

STATISTICAL ANALYSIS

Comparisons between delinquent and control groups was accomplished through application of the conventional formula for the critical ratio, $\frac{D}{PE\ (diff)}$, where D is the difference between mean differential weighted scores of similar and mixed life age groups and PE (diff) is the probable error of difference between the same means.

Based on frequency distributions of differential weighted scores, Table 1 shows (a) comparative quartiles, standard deviations, medians, and means for four life age groups of delinquents and non-delinquents; and (b) differences of means between control and experimental groups with the resulting critical ratios.

Table 1

Comparison of Delinquents and Non-Delinquents in Differential Weighted Scores on Interest-Attitude Tests

Group	No. of Cases	Q_1	Q_3	SD	Median	Mean	Diff. of Means	$\frac{D}{PE(diff)}$
14-yr. delinquent	66	13.2	24.6	7.76	19.6	19.3		
14-yr. non-delinquent	76	3.7	13.3	7.04	8.9	9.1	10.2	12.0
15-yr. delinquent	75	11.4	24.0	8.74	16.9	18.2		
15-yr. non-delinquent	101	4.5	13.0	6.77	8.3	9.3	8.9	10.9
16-yr. delinquent	112	9.4	21.6	7.97	15.4	15.9		
16-yr. non-delinquent	100	2.2	10.0	5.79	5.8	6.7	9.2	14.4
17-yr. delinquent	63	9.5	19.8	7.65	15.1	15.3		
17-yr. non-delinquent	97	1.3	9.2	6.68	4.2	5.9	9.4	11.8

Since $\frac{D}{PE(diff)}$ must be at least 4 to insure complete reliability, it is evident that all of the differences between means shown in Table 1 are significant, each critical ratio being well in excess of the magnitude required.

Another interesting aspect of the comparative groups is seen in terms of overlapping. Considering the life age groups from the point of view of median differential scores for delinquents, approximately 94 per cent of 14-year non-delinquents, 88 per cent of 15-year, 91 per cent of 16-year, and 92 per cent of 17-year fall <u>below</u> the median of the same life age groupings of delinquents. Or, stated differently, about 6, 12, 9, and 8 per cent of the respective life age groups, 14, 15, 16, and 17 year non-delinquents, <u>exceed</u> the medians of similar life age groupings of delinquents.

By way of more complete verification of the differences in personality characteristics between delinquents and non-delinquents the technique of differential scoring has been applied to two entirely new samples. One of these consists of a group of 13-year delinquents compared with a control group of similar life age.

The other consists of 115 delinquents of heterogenous life age, the median being 16.6 years. The latter sample was obtained practically five years after the original life age groups on the basis of which differential items were selected. As a control group for the latter, the differential scores of 374 non-delinquent subjects cited in Table 1 were pooled in one frequency distribution. Results of the last comparison are seen in Table 2.

Table 2

Comparison of Delinquents and Non-Delinquents in Differential Weighted Scores on Interest-Attitude Tests, Based on New Samples

Group	No. of Cases	Q_1	Q_3	SD	Median	Mean	Diff. of Means	$\frac{D}{PE(diff)}$
13-yr. delinquent	34	11.9	25.6	9.17	18.8	19.6		
13-yr. non-delinquent	61	4.0	14.9	7.97	9.1	10.5	9.1	7.2
Mixed ages, delinquent	115	10.0	20.7	7.47	14.8	15.3		
Mixed ages, non-delinquent	374	2.6	11.7	6.72	6.7	7.7	7.6	14.5

Differences between means shown in Table 2 are entirely reliable. The differences between control and experimental groups of mixed life ages are especially worthy of note. In terms of overlapping about 84 per cent of 13-year non-delinquents fall below the median differential score for delinquents, or 16 per cent exceed that median. In the mixed age group approximately 87 per cent of the control subjects are below the median for delinquents, or 13 per cent fall above that point.

It may be concluded from the foregoing analysis that a significant difference in personality exists between delinquents and normally adjusted subjects. That is, the constituent traits studied appear with a frequency among delinquents which reliably differentiate them from non-delinquents.

As a means of expressing an individual's emotional development Pressey (3) has proposed the concept emotional age. Emotional age may be derived from total scores on the Interest-Attitude Tests. In another connection the present writer (2) found that delinquents show on the average about two and one-half years emotional retardation. A possibility suggested itself that a relationship might exist between emotional ages and differential weighted scores of the juvenile delinquents. Hence, Pearson r's have been computed between the variables mentioned based on four life age groupings of delinquents appearing in Table 1. The resulting correlations appear in Table 3.

Table 3

Correlations between Emotional Ages and Differential Weighted Scores

Group	No. of Cases	r	PE
14-year delinquents	66	-.69	±.04
15-year delinquents	75	-.77	±.03
16-year delinquents	112	-.71	±.03
17-year delinquents	63	-.62	±.05

On the Interest-Attitude Tests the higher the score, based on the total examination, the less mature emotionally is the subject. For instance, a total score of 213 results in an emotional age of 10 years, while a total score of 14 gives

an emotional age of 18 years. Quite manifest, from the series of correlations in Table 3, is the fact that there is a highly significant and reasonably consistent negative relationship between emotional age and differential weighted score. This means that a tendency is present toward varying concomitance between emotional immaturity and differential scores of delinquents. Stated in other terms, subjects who have the highest scores on the Interest-Attitude Tests tend likewise to respond more often in terms of differential items.

A final aspect of the present investigation has to do with the relationship between differential weighted scores and degree of delinquency. The writer (1) has devised a method by which variations in the seriousness of delinquent careers may be estimated. Degree of delinquency is stated in terms of a delinquency index. As will be noted from Table 4 little meaning can be attached to correlations between differential weighted scores and degree of delinquency. Subjects used are the life age groups of delinquents found in Table 1 and later in Table 3

Table 4

Correlations between Differential Weighted Scores and Delinquency Indexes

Group	No. of Cases	r	PE
14 year delinquents	66	-.09	± .08
15-year delinquents	75	-.15	± .08
16-year delinquents	112	.13	± .06
17-year delinquents	63	.11	± .08

Although two of the foregoing correlations are negative and two positive, one characteristic in common is their insignificance. It is evident that two almost entirely different aspects of delinquent personality are measured by the devices the results of which have been correlated.

PRACTICAL CONSIDERATIONS

One indubitable fact may be observed from this study: within the limits of the traits differentiated, delinquents manifest a palpable difference from normally adjusted individuals. In relation to those circumstances which are considered wrong, cause worry or anxiety, elicit interest, or traits which are admired in others the delinquent presents an atypical picture. Of course, no contention is made that delinquents and non-delinquents fall into mutually exclusive groups in the differential traits. As has been pointed out there is some overlapping. The results suggest strongly, however, the possibility of employing the Interest-Attitude Tests as one means of anticipating delinquent conduct or apprehending delinquent tendencies which have not become overt. In the groups examined from 6 to 16 per cent of the control groups exceeded the median differential scores of respective delinquent groups. All of which may be interpreted as meaning that clinically one would be warranted in suspecting potential delinquency when an individual's score on the differential items exceeds a certain value. It is impossible to state specifically the magnitude of this crucial value. Medians of the six delinquent groups studied range approximately from 15 to 20. Interpreted conservatively it is likely that a score on the differential items of fifteen or more is sufficient evidence for an intensive study and follow-up of the individual case, even though no obvious delinquency has been recorded.

A fact however, which should not be overlooked is that differential scores are a function somewhat of higher total scores on the Interest-Attitude Tests. This brings retarded emotional development into the picture. Hence, two criteria are available in the differential diagnosis of suspected juvenile delinquency, viz., emotional age and differential weighted score. That is, if an individual is two and one-half or more years retarded in emotional age as compared with his life age and at the same time scores significantly high on the differential items, there is every reason for scrutinizing his behavior closely for possible delinquent tendencies.

From the standpoint of interpreting results a possible limitation of this investigation should be mentioned. All of the various experimental groups were subjects confined in an institution for juvenile delinquents. Whether or not differential items reflect certain attitudes, prejudices, and tendencies engendered by institutionalization cannot be stated definitely on the basis of the data at hand. An answer to the question will depend upon further experimentation, especially application of the method to non-institutionalized delinquent subjects and behavior problem cases in the public schools.

SUMMARY

Section two of a broader investigation of the personality characteristics of juvenile delinquents has been an attempt to test the reliability of a group of differential items methodologically ascertained in section one. That is, the primary purpose of the present section has been to discover to what extent the differential items were diagnostic of actual differences in personality between delinquent and non-delinquent subjects. In all instances in which comparisons have been made, significant differences have been demonstrated. Two relationships of the differential items have been found, one significant, the other without significance. In the first case it has been shown that differential weighted scores of delinquent subjects are related to emotional age. Secondly, it has been found that little or no relationship exists between differential scores and degree of delinquency. Some suggestions have been indicated as to possible ways of employing the Interest-Attitude Tests in the clinical analysis of delinquent behavior.

REFERENCES

(1) Durea, M. A.: A quantitative method for diagnosing the seriousness of asocial behavior of juvenile delinquents. J. Gen. Psychol., 1936, 14, 412-421.

(2) Durea, M. A.: The emotional maturity of juvenile delinquents. J. Abnorm. and Social Psychol., 1937, XXXI, No. 4.

(3) Durea, M. A.: Personality characteristics of juvenile delinquents. I. A Method for the selection of differentiating traits. Child Development, 1937, 8, 115-128.

(4) Pressey, S. L., and L. C.: Development of the interest-attitude test. J. Appl. Psychol., 1933, 17, 1-16.

A PRELIMINARY REPORT OF DARK ADAPTATION IN YOUNG INFANTS

JANET E. REDFIELD [1]

This article presents some preliminary findings of a study on the problem of dark adaptation of the infant eye.

Since the infant's responses to stimuli are characterized by much variability, the amount of bodily activity was employed as a criterion in a number of sensory studies by Irwin (1, 2), Weiss (9), Stubbs (8), Smith (7), and Richards (5, 6). It is the index adopted in the present investigation. Irwin and Weiss (3) and Smith (7) also studied the influence of visual stimulation upon crying. Accordingly a second measure, amount of crying, has been used as a possible index of dark adaptation.

Weiss (9) found that increasing the intensity of light resulted in a decrease in bodily activity. This result has been corroborated by Irwin and Weiss (4) and by Richards (5, 6). The hypothesis may be suggested that if the sensitivity of the infant's eyes could be increased and tested by a light of constant intensity, the same result found by the above workers should be obtained. Their method was to vary the light intensity while holding dark adaptation constant; here dark adaptation is varied and intensity is held constant. Theoretically, as the period of dark adaptation is lengthened the amount of body activity should decrease. The present study was undertaken to test this hypothesis.

The infant[2] was brought from the hospital nursery and placed on the stabilimeter within a Pratt experimental cabinet. All clothing except the diaper was removed, and the infant was left in the cabinet while the stabilimeter and recording apparatus were adjusted before beginning the experiment. Thus an interval of about five minutes was allowed to insure that the influence of handling might not affect the responses.

Each experiment consisted of three parts: (1) one minute of an adapting light, 47 foot candles, to insure uniform initial light adaptation of the eye; (2) one of four experimental periods of darkness which consisted of one, five, ten, and twenty minutes; and (3) five minutes of the control light, 1.5 foot candles. The adapting light was supplied by a 100 watt white frosted bulb, and the control light by a 7.5 watt frosted bulb.

All illumination was excluded from the cabinet during the dark period. The only light in the laboratory during the experiment was a small shaded bulb above the control board, supplying sufficient light for the experimenter to time and record observations. Experiments were performed during the last half of the interfeeding period when activity may be expected to be greatest.

Forty-seven infants ranging in age from seven to nine days were given 200 trials, fifty in each of four experimental periods. Twenty-three of the subjects were boys and twenty-four girls; twenty-five trials were made at each interval with each sex. The measure of response used was the number of active seconds per minute during the control light period following each interval of darkness; it

[1] From Iowa Child Welfare Research Station, State University of Iowa, Iowa City, Iowa.

[2] The subjects were available through the courtesy of Dr. E. D. Plass, head of the Department of Obstetrics and Gynecology in the hospital of the State University of Iowa.

was recorded by means of an electric stabilimeter and polygraph. In addition quarter-minute observations were made of eyes open or shut and whether the infant was silent or crying. The per cents of agreement in counting activity records between the experimenter and two assistants were 97 and 99 respectively; 100 per cent agreement was obtained for observations of condition.

RESULTS

The data[1] have been analyzed (1) for the group as a whole, (2) for sex differences, (3) for crying and silent subjects, and (4) for infants with eyes open and eyes closed.

The Group as a Whole

Mean Activity.- Mean activity for the group as a whole following each of the experimental intervals, together with the standard deviation of each distribution and the standard errors of the means, are found in the following tabulation:

Adaptation Time (Minutes)	Mean	Standard Deviation of Distribution	Standard Error of Mean
1	16.6	18.1	2.6
5	13.6	15.8	2.2
10	12.2	14.2	2.0
20	9.4	12.5	1.8

It will be seen that bodily activity decreases as the time of dark adaptation is lengthened (Figure 1). The significance of the differences between activity in each of the periods following dark adaptation is as follows:

Adaptation Time (Minutes)	Ratio of Difference to Standard Error of Difference	Chances in 100 Difference Is Greater Than Zero
1 and 20	2.31	99
5 and 20	1.48	93
10 and 20	1.03	85
1 and 10	1.35	91
5 and 10	.47	69
1 and 5	.88	82

While the differences in mean activity following these intervals of dark adaptation are not statistically significant, the difference is always in the same direction. The largest value for the critical ratio is 2.31. It occurs in the case of activity following dark adaptation periods of one and twenty minutes, which indicates that if the experiment were repeated under like conditions the chances are 99 in 100 that the difference in bodily activity would be as great.

Change in Activity Between Long and Short Intervals.- Another method of analyzing the data is to determine the per cents of infants showing increases, decreases, and no differences in activity in short versus long intervals. The results of this analysis for the cases in which the same infant served as subject for both a long and a short period are shown below:

	Adaptation Time, Minutes	
	1 and 20*	5 and 20**
Per cent of cases showing increase	37	25
Mean increase of activity	11.4	3.1
Per cent of cases showing decrease	50	40
Mean decrease of activity	20.5	17.8
Per cent of cases showing no change	13	35

*24 cases **20 cases

[1] The original data upon which this report is based may be found in the appendix of the manuscript copy on file in the University of Iowa Library.

It is apparent from this tabulation that the mean decrease in both comparisons is several times the increase, although the per cent of infants showing a decrease in activity is low.

Percentiles.- An analysis of the data by percentiles has also been made. The tabulation shows that the 90th, 75th, and 60th percentiles give the same trend as revealed by the means; that is, that activity decreased with increase in the length of the dark adaptation period. The two lower percentiles, however, are not consistent, probably because some infants exhibited little activity.

Percentiles	Dark Adaptation Time, Minutes			
	1	5	10	20
90	43.5	40.5	37.5	30.0
75	33.8	27.0	17.0	13.0
60	17.5	13.5	11.5	5.7
50	10.5	6.5	7.5	3.2
25	.7	.5	1.2	.3

Immediate Response to Change: The First Minute.- Two analyses of response to change from darkness to light have been attempted. The frequency of change in general level of activity from the last minute of darkness to the first minute of light for each of the experimental intervals is shown below:

Change in Activity	Time in Darkness, Minutes			
	1	5	10	20
Increase	3	2	4	1
Decrease	8	12	13	11
No change	39	36	33	38

It is evident from this analysis that the immediate response cannot be used as a measure of dark adaptation when change in activity is the criterion.

Mean activity in the first minute of light after darkness is the second measure used. The means for the four intervals are respectively 17.4, 13.7, 11.3, and 8.1. The critical ratio of the means after one and twenty minutes in darkness is 2.62, indicating that there are 99 chances in 100 for a difference between the true means. The coefficients of correlation between these results and the means of the five-minute period are .82 $\pm$ .05, .48 $\pm$ .11, .88 $\pm$ .03, and .85 $\pm$ .04.

Per Cent of Crying.- Smith (7) found that per cent of crying was a reliable measure of the response of infants to the brightness of three hues. The frequencies of crying after each period of dark adaptation in the present experiment are 35.3, 30.7, 33.4, and 24.6 per cent respectively. Thus while there is a reduction in amount of crying from the shortest to the longest interval, the decrease is not progressive.

Sex Differences

For purposes of comparison, the means in activity for boys, for girls, and for all cases are shown in the following tabulation and in Figure 1:

Adaptation Time (Minutes)	Boys	Girls	All Cases
1	16.7	16.6	16.6
5	15.8	11.5	13.6
10	12.3	12.2	12.2
20	8.7	10.2	9.4

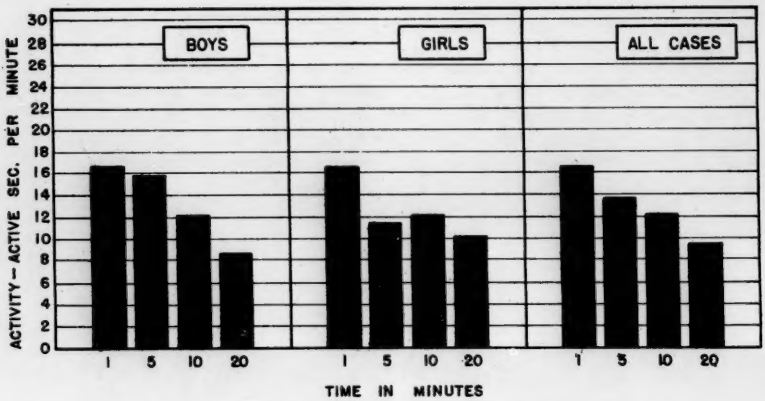


Figure 1. Mean Activity for Boys, for Girls, and for all Subjects.

It will be seen that differences in activity between the sexes are negligible after one and ten minutes of darkness; after five minutes the boys are more active, and after twenty minutes the girls.

The per cents of crying for the two sexes are given below:

Adaptation Time (Minutes)	Boys	Girls
1	38.2	32.4
5	36.2	25.2
10	39.6	27.2
20	23.6	25.6

The boys consistently cry a greater part of the time, except after twenty minutes of darkness.

Crying and Silent Subjects

The effect of a distinction between crying and silent subjects is to eliminate those who were asleep during the whole or greater part of the experiment. Mean activity for these groups in the experimental intervals, together with the number of subjects in each case, is given below:

Adaptation Time (Minutes)	Infants			
	Crying		Silent	
	Number	Mean	Number	Mean
1	28	28.36	22	1.68
5	29	23.12	21	.53
10	33	18.03	17	.95
20	26	17.62	24	.56

Infants With Eyes Open and Those With Eyes Closed

An analysis of those infants whose eyes were open 50 per cent or more of the time and those whose eyes were closed 50 per cent or more of the time should evaluate the importance of these factors. However, because of the small number of

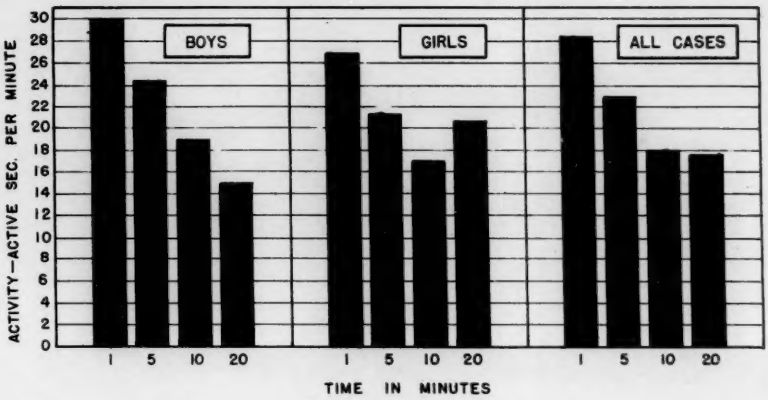


Figure 2. Mean Activity for Crying Subjects

cases in which the eyes were open, no detailed statistical procedure could be used. The means in activity and the numbers of infants showing these conditions are given below:

Adaptation Time (Minutes)	Eyes Open		Eyes Closed	
	Number	Mean	Number	Mean
1	10	14.8	43	16.8
5	7	8.6	44	14.5
10	12	11.1	40	12.4
20	7	9.7	44	9.4

From these data it is seen that the infant whose eyes are closed gives results consistent with findings for the whole group. It thus appears that activity is inhibited following increased duration of darkness even though the eyes are closed.

SUMMARY

The following tabulation summarizes the criteria which give consistent results in measuring the responses of infants to a constant intensity of light after varying degrees of dark adaptation:

Criterion	Adaptation Time, Minutes			
	1	5	10	20
Mean activity, 5 minutes	16.6	13.6	12.2	9.4
Mean activity, 1 minute	17.4	13.7	11.3	8.1
Mean activity, crying subjects	28.4	23.1	18.0	17.6
90th Percentile	43.5	40.5	37.5	30.0
75th Percentile	33.8	27.0	17.0	13.0
60th Percentile	17.5	13.5	11.5	5.7

This tabulation summarizes the criteria which give inconclusive results:

Criterion	Adaptation Time, Minutes			
	1	5	10	20
50th Percentile	10.5	6.5	7.5	3.2
25th Percentile	.7	.5	1.2	.3
Frequency of immediate decrease in activity in light after darkness	8.0	12.0	13.0	11.0
Per cent of crying subjects	35.3	30.7	33.4	24.6
Mean activity, silent subjects	1.7	.5	1.0	.6

These results may be summarized as follows:

1. Mean activity - whether over a five-minute period or for one minute - in constant light after dark adaptation decreases with lengthened periods of darkness from one to twenty minutes. The coefficients of correlation between these two measures after one, five, ten, and twenty minutes in darkness are respectively .82 ± .05, .48 ± .11, .88 ± .05, and .85 ± .04.

2. On the basis of bodily activity, the crying infant shows a more consistent response to dark adaptation than does the silent one.

3. The 90th, 75th, and 60th percentiles show consistent results; this is not true of the lower percentile ranks, probably because of the large numbers of inactive or slightly active infants.

4. The frequency of immediate response to light fails as a measure of dark adaptation.

5. Per cent of crying in light decreases with lengthened dark adaptation, except following the ten minute period.

6. The responses of infants with eyes closed show a high degree of consistency with the results for all subjects; the inconsistent results given by infants with eyes open are probably explained by the small number of the cases in which this condition was found.

7. The boys are slightly more active than the girls, except following the twenty-minute period; they are also less variable.

8. The boys cry a greater per cent of the time than do the girls, except following the twenty-minute period.

The present study presents evidence concerning the adequacy of the criteria used for measuring the dark adaptation of the infant eye. It has been shown that bodily activity decreases following lengthened periods of dark adaptation, but in no case has statistical significance been obtained between activity levels. However, there are 99 chances in 100 that the difference between the results of the shortest and longest dark adaptation periods is a true difference. It follows that this measure is adequate to detect differences in the sensitivity of the infant eye after dark adaptation.

REFERENCES

(1) Irwin, Orvis C.: The amount and nature of activities of newborn infants under constant external stimulating conditions during the first ten days of life. Genet. Psychol. Monog., 1930, 8, 1-92.

(2) Irwin, Orvis C.: The distribution of the amount of motility in young infants between two nursing periods. J. Comp. Psychol., 1932, 14, 429-445.

(3) Irwin, Orvis C., and Weiss, LaBerta A.: Differential variations in the activity and crying of the newborn infant under different intensities of light: A comparison of observational with polygraph findings. (In) Irwin,

Orvis C., Weiss, LaBerta A., and Stubbs, Esther M.: Studies in Infant Behavior I. Univ. Iowa Stud., Stud. in Child Welfare, 1934, 9, No. 4, Pp. 175. (p. 137-162)

(4) Irwin, Orvis C., and Weiss, LaBerta A.: The effect of darkness on the activity of newborn infants. (In) Irwin, Orvis C., Weiss, LaBerta A., and Stubbs, Esther M.: Studies in Infant Behavior I. Univ. Iowa Stud., Stud. in Child Welfare, 1934, 9, No. 4, Pp. 175. (p. 163-175)

(5) Richards, T. W.: The relationship between bodily and gastric activity of newborn infants. I. Correlation and influence of time since feeding. Human Biology, 1936, 8, 368-380.

(6) Richards, T. W.: The relationship between bodily and gastric activity of newborn infants. II. Simultaneous variations in the bodily and gastric activity of newborn infants under long-continued light stimulation. Human Biology, 1936, 8, 381-386.

(7) Smith, Josephine M.: The relative brightness values of three hues for newborn infants. (In) Wenger, M. A., Smith, Josephine M., Hazard, Charles, and Irwin, Orvis C.: Studies in Infant Behavior III. Univ. Iowa Stud., Stud. in Child Welfare, 1936, 12, No. 1, Pp. 207. (p. 91-140)

(8) Stubbs, Esther M.: The effect of the factors of duration, intensity, and pitch of sound stimuli on the responses of newborn infants. (In) Irwin, Orvis C., Weiss, LaBerta A., and Stubbs, Esther M.: Studies in Infant Behavior I. Univ. Iowa Stud., Stud. in Child Welfare, 1934, 9, No. 4, Pp. 175. (p. 75-135)

(9) Weiss, LaBerta A.: Differential variations in the amount of activity of newborn infants under continuous light and sound stimulation. (In) Irwin, Orvis C., Weiss, LaBerta A., and Stubbs, Esther M.: Studies in Infant Behavior I. Univ. Iowa Stud., Stud. in Child Welfare, 1934, 9, No. 4, Pp. 175. (p.9-74)

DEVELOPMENT IN THE USE OF PROPER NAMES

EDITH A. DAVIS [1]

The proper names used by 436 children in 21,800 remarks were recorded under a standardized situation [2] and were found to show certain well defined age differences. Subjects were selected at three discrete age levels, 5 1/2, 6 1/2, and 9 1/2 years, on a percentage basis that was representative of the Minneapolis-St. Paul population, the father's occupation being the criterion. The grade location of subjects was for the most part kindergarten, first grade, and fourth grade. Individual interviews during school hours lasted in each case until 50 remarks were recorded.

With two exceptions, the numerical importance of proper names in children's language proved to be slight. The subjects were given an opportunity to play with a collection of Wild West toys. Probably because of this fact the word Indian occurred, 1,656 times, and ranked fifteenth in the list of all words used. Christmas was noted 86 times. These two words, as well as other names of months, days, and holidays; words describing nationality, and names of deity, were listed in the regular manner, and the category of proper names was reserved for names of specific persons and places. Aside from the terms Indian and Christmas, names in the above categories were of little importance in the final tabulations, since only 32 expressions were involved. The inclusion of Indian and Christmas, however, would have given so distorted a picture that the decision to limit the category to persons and places seems a wise one.

The number of proper names used per child was slightly less at 9 1/2 years than at the earlier levels, both absolutely and when related to the total words used per child, as is shown in Table I.

Table I

Mean Number of Proper Names

Age in years	Number of Subjects	Total number proper names	Mean per child	Per cent of all words per child	Total number different proper names	Mean per Child	Per cent of mean number different words
5 1/2	248	268	1.08	0.5	197	0.79	0.8
6 1/2	63	87	1.38	0.5	56	0.89	0.8
9 1/2	125	111	0.89	0.3	93	0.74	0.6

More detailed analysis indicated that the persons named by the young child tend to be friends, teachers, and members of the family, while the older child mentions historical and fictitious characters. Table II shows the distribution of surnames by age.

A similar name trend appeared in the analysis of places mentioned. The young child refers to local places - schools, parks, his father's place of business, his own street or telephone number - while the older child mentions places at a

[1] From the Institute of Child Welfare, University of Minnesota.

[2] Davis, Edith A. The Development of Linguistic Skill in Twins, Singletons with Siblings, and Only Children. University of Minnesota, Institute of Child Welfare Monograph Series Number XIV. In Press.

Table II

Distribution of Surnames Used

Age in Years	Friends	Teachers	Members of Family	Total	Mean Per Child
5 1/2	24	15	1	40	.16
6 1/2	2	2	1	5	.08
9 1/2	7	0	1	8	.06

distance, other states, and foreign countries. At the kindergarten age the child is still fascinated by names as names, and the distinction of individuals by this means remains of major importance. By 9 1/2 years all this has become an old story, but there is a delight in discussing Hiawatha, California, and Holland.

The 248 5 1/2 year old children used 69 different Christian names, which is a ratio of one name to each three children. This ratio declines to one name for each five children at 9 1/2 years. Table III shows this relation and in addition gives the ratios for the number of different names used, i.e. with repetitions of the same name excluded.

Table III

Mean Number of Christian Names

Age in Years	Total Names	Mean per Child	Number of Different Names	Mean per Child
5 1/2	75	.30	69	.28
6 1/2	20	.32	9	.14
9 1/2	24	.19	21	.17

Thirteen of the Christian names occurred three times or oftener. All except one of these (Elaine) appears in Thorndike's list of 20,000 words. Leonard and Jimmie are found in Thorndike's sixth thousand, the other ten names in the first five thousand. The names, with their frequencies, are given in Table IV.

Table IV

Distribution of Christian Names

Name	Times used	Thorndike
Billy	12	3b
Betty	9	4a
Jimmie	9	6
Tom, Tommy	5	3a, 5a
George	6	2a
Dick	4	2b
John	4	1a
Jack	3	2a
Mary	4	2a
Dorothy	3	3b
Walter	3	3b
Leonard	3	6
Elaine	3	-

George Washington and Buffalo Bill were the only historical personages mentioned by the 5 1/2 year olds. Santa Claus, Little Black Sambo, Tarzan, Wolfpaw, and Little Bo Peep were the fictitious characters to whom they alluded. At 9 1/2 years eight historical and nine fictitious characters were mentioned. The distribution of the reference to such characters is shown in Table V.

Table V

Distribution of Historical and Fictitious Characters

Historical	Ages 5 1/2	Ages 6 1/2	Ages 9 1/2	All Ages
George Washington	9	3	1	13
Buffalo Bill	3		4	7
Kit Carson			2	2
Custer			2	2
Abraham Lincoln, Mohicans, Radisson, Groseillers (1 each)				4
Fictitious				
Santa Claus	28	5		33
Little Black Sambo	13	21	3	37
Tarzan	2			2
Wolfpaw	2		1	3
Bo Peep	1			1
Tom Sawyer		3		3
Hiawatha			3	3
"Frederick"			2	2
Redwing, Jack Frost, Old Mother Hubbard, "Billy" (1 each)				4
Great Spirit, "Polly", Buck Jones, Peter Rabbit, Alice in Wonderland (1 each)				5

We may conclude that as a child develops he is less likely to name the persons and places associated with his every day life. His casual conversation, however, does reflect his increasing familiarity with faraway places and with historical and fictitious characters.

MOTIVATING FACTORS IN CHILD LEARNING

ELISABETH T. MAST[1]

There are many problems of motivation with young children which need to be investigated. There are few studies on the influence of incentives on young children and the motivating factors in the experiments on learning in young children give few data for evaluation of incentives and rewards. The present experiment was run as the second series in an experiment by Johnson.[2] It was planned to test the influence of an intrinsic reward plus an unpleasant incentive on the speed of learning to open a simple problem box.

METHOD

The method was in general experimental. The learning problem was presented to each subject individually. A period of five minutes was chosen as the maximum length of a single learning period, since it is difficult to get the child's attention to one problem for a long period. Learning periods were given at intervals of two to three days until the problem was mastered. A retest was given approximately one week after the first success. Five of the subjects were given a retest seven months later.

SUBJECTS

The subjects were forty-three children enrolled in the Johns Hopkins University Child Institute during nineteen thirty-six and nineteen thirty-seven. Two of the children did not complete the test due to withdrawal from school and three refused to cooperate on the test. For convenience in analysing the data the subjects are considered in two groups. Group A was composed of twenty-three children who were tested and retested in May nineteen thirty-six; Group B was composed of twenty children tested in December nineteen thirty-six. The range in chronological age of the subjects in Group A was from thirty months to seventy months with an average of forty-five and nine-tenths months; the range of those in Group B was from twenty-two months to fifty-nine months with an average age of forty months. All ages were taken from the date of the first presentation.

The range in I.Q. as indicated by the Minnesota Preschool Scale was from seventy-two to one hundred and forty-one with an average of one hundred and nineteen for Group A and from fifty to one hundred and twenty-eight with an average of one hundred and sixteen for Group B. These scores are not assumed to be reliable indices of the mental abilities since some children give poor attention or cooperation on the test. Intelligence scores for young children can do no more than act as an indication of the general level of intelligence of the children. The subjects came from average and above average homes as rated by economic status.

[1] From The Child Institute, Johns Hopkins University.

[2] Buford Johnson. Variations in emotional responses of children. Child Development, 1936, 7, 2.

APPARATUS

The apparatus consisted of a wooden box, nine inches by five inches by five inches, with a glass top hinged at the back. The box could be opened by inserting a stylus in a quarter inch hole in the center of a one inch brass plate on the front of the box which released the catch holding down the top of the box. An illustration of the box appears in an article by Johnson. The apparatus was so wired that an ordinary doorbell rang when the stylus came in contact with the plate.

The problem box was placed on a low white table with the stylus placed to the right of the box. A small kindergarten chair for the subject was put directly in front of the table. A chair for the experimenter was placed to the left and slightly behind the subject's chair.

The reward incentives were placed inside the problem box and were visible through the top of the box. The rewards used were small toy rubber autos from F. W. Woolworth's. All of the cars used in the first presentation were red and of the same model. Those used in the retest were green or blue and were of different models.

TESTING PROCEDURE

The children were obtained from the group by the experimenter and taken individually to the experimental room. After the child had been seated, the experimenter said, "Do you see what is in this box? When you have opened the box you may have it. You must take this in your hand (the experimenter handed the stylus to the child) and open the box with it." Time with a stop-watch was taken from the moment the directions were completed until the box was opened or until five minutes had elapsed. If the child failed to open the box in five minutes, the experimenter suggested that the child come back another day to open the box. Running notes of the speech of the child and of his overt emotional behavior during the testing period were taken. If the child attempted to open the box with his hands the experimenter said, "No, you must open it with this (the stylus)." The same directions were used on succeeding trials except that the experimenter said, "You know what is in this box, don't you?, etc." Questions of the subjects during the testing period were answered briefly and without giving any clue to the solution of the problem. To the numerous requests for the experimenter to open the box, she replied, "I know you can do it yourself. When you have opened the box then you may have the car to take home." Encouragement was given whenever it seemed advisable.

The subjects in Group A were given a retest several days after success on the first test. At the close of this test the experimenter asked, 'How do you open the box?" Verbal replies and the method of demonstration were recorded. No retest was given to the subjects in Group B since the time scores for the retest of Group A showed that the problem was thoroughly mastered during the first trial and also since many of the time scores on the first test for Group B were so small that no improvement would be possible. The subjects in this group were therefore asked how to open the box after the first success.

ANALYSIS OF EXPERIMENTAL RESULTS

The results of this investigation have been studied: (1) relative to age and

I.Q. as a factor in the speed of mastering the problem; (2) the emotional factors involved; (3) the relative influence of the motivating factors employed; (4) a comparison of the results of Group A and of Group B.

THE INFLUENCE OF AGE AND I.Q.

The time scores, age and I.Q. of the subjects in Group A are presented in Table I; those of Group B are given in Table II. An examination of these Tables indicates that there is no relation between either age or I.Q. and the time score. In Group A the youngest and the oldest child learned the problem with almost equal rapidity. The highest and the lowest scores were distributed throughout the age range. The situation in Group B was found to be similar. Of the two children in Group A who obtained the highest scores on the intelligence test, one had one of the smallest time scores and the other had the largest. The child who had the smallest time score also had the lowest I.Q. Although the I.Q. scores for Group B are incomplete, no correlation between the I.Q. and the speed of learning is evident.

EMOTIONAL FACTORS

From a study of the verbal remarks of the subjects in relation to the time scores presented in Table I and II, there appears to be some correlation between the emotional factors involved and the speed of mastering the problem. In nearly all of the instances in which the time scores were large, the subjects showed emotional disturbance and usually made some verbal comment indicating a dislike for the sound of the bell. Only three children in Group A and one in Group B stated that they enjoyed the sound of the bell and wished to ring it just for the sake of the noise. Eight children in Group A and three in Group B expressed a definite dislike for the noise. Johnson had analysed the emotional factors involved in the learning of this problem by the subjects in Group A. There appeared to be less emotional disturbance in Group B than in Group A.

All but two of the subjects in Group A reduced their time on the second trial. W. G.'s score on the sound performance was only slightly higher and it is probable that she opened the box by mistake the first time. T. B.'s scores were almost identical. The speed of opening the box on the second trial does not appear to depend on age or I.Q. Since no unusual nor significant results were obtained on the second presentation with Group A and since the time scores obtained by Group B were already very small, only one trial was given to the subjects in Group B.

The results of the five subjects retested after seven months are given in Table III. H. L., who had refused to do the experiment after the first sound of the bell, continued to be non-cooperative. Of the other four, two of the children reduced their time scores and the other two increased their time scores. The two who reduced their time scores had asked for "the game with the box and the autos" several months before the retest was given. The subjects who took longer to open the box appeared to remember only a part of the principle involved so had to unlearn this before they could relearn the correct principle.

Some of the children had apparently ascertained the principle of the box before they opened it. P. T. said, "I don't want the bell to ring when I put it (the stylus) against there (the plate). Where else shall I put it to open the box?" After three hundred seconds S. A. said, "I have to ring it before it opens" but

Table I

Subject	Time score in sec. Trial 1	Trial 2	Age in months	I.Q.	Emotional disturbance as indicated by overt behavior or verbal comments.
T.B.	32	35	36	72	None
Y.B.	65	21	36	123	None
S.G.	67	2	64	135	None
W.C.	70	167	47	108	Slight
M.K.	110	2	70	114	None
W.E.	115	2	30	141	None
S.L.	120	20	53	126	Slight
L.M.	200	155	37	108	None
S.F.	215	145	53	126	None
M.M.	240	4	36	130	None
T.J.	305	4	64	137	None
T.M.	345	176	44	117	"I don't like that noise."
Z.J.	366	2	48	117	"I don't like that noise."
K.J.	452	9	31	118	"Bell hurts my ears."
H.K.	613	212	37	111	"That big noise scares me to death."
H.R.	860	89	37	102	None. Was fascinated by the sound.
L.H.	917	4	49	121	Jumped each time bell rang.
S.A.	947	94	36	119	"May I ring?" Fascinated by the bell.
K.N.	1130	182	69	193#	Showed a definite dislike for bell.
D.D.	1204	190	53	103	Jumped each time the bell rang.
J.P.	1498	6	38	141	Slight, Experimented with methods of ringing bell.
M.W.	refused		38	119	As soon as the bell rang said, "No I won't touch it."
H.L.	refused		49	126	As soon as bell rang darted out of the room and refused to return.

- Measured on the Stanford Binet scale.

Table II

Subject	Time score in sec.	Age in months	I.Q.	Emotional disturbance as indicated by overt behavior or verbal comments.
H.C.	2	38		None
P.T.	2	37	105	None
K.H.	2	38	112	None
H.N.	3	22	96	Screamed when the bell rang but had already opened the box.
L.P.	3	35	102	None
M.R.	3	47	131	None
S.H.	6	32	95	None
B.S.	7	46		None
M.B.	8	37	50	None
B.J.	38	48	130*	Slight
M.S.	40	44		None
M.E.	60	42	128	None
F.B.	70	24		None
P.J.	393	29		None. Was fascinated by the sound.
Z.V.	591	49		None
B.F.	935	59	126*	Deep breathing and facial expression showed emotional disturbance.
B.M.	not completed	39	128*	"It scared me." "Where is the hole you put it in?"
P.T.	not completed	56		"I don't want the bell to ring."
D.T.	refused	34	126	"If I put this in it will ring." Refused to cooperate but opened the box in ten seconds after the bell was discontinued.

* - Measured on the Stutsman Test.

she took nine hundred and forty-seven seconds to open the box. H. N., after one hundred and eighty seconds, said, "You open it down here but it's going to ring." However, none of those children distinguished between the sound of the bell and pushing with the stylus as the effective factor in opening the box. T. J. was the only child to do this. She said on her retest, "The bell doesn't open the box, does it?" Several of the children gave as an explanation for the method of opening the box, "You ring down here." On her retest seven months later, T. N., said, "It looks like it's coming open every time I ring the bell. ...I think if

Table III

Subject	Time score in sec.	Age in months	I.Q.	Emotional disturbance as indicated by overt behavior or verbal comments.
H.R.	3	44	102	Made no remarks but opened the box immediately.
J.P.	3	45	141	Before she entered the room, she told how to open the box.
T.M.	403	51	117	Disturbed by the noise. Said, "It looks like it's coming open every time I ring the bell."
Y.B.	780	43	123	No disturbance but seemed to have forgotten a part of the principle. Kept stylus on the plate most of the time.
H.L.	refused	56	121	She suggested coming to the experimental room but would not enter the room. She stood in the doorway and talked about the unpleasantness of the bell.

I make a big noise it might get out or pop out." It was difficult to determine whether children actually understood the principle since many of them gave no verbal clue as to their comprehension, the majority of them merely demonstrating how they had opened the box when they were asked for the principle.

INFLUENCE OF THE MOTIVATING FACTORS

The emotional influence of the bell has already been discussed. The comments of the children indicated that the autos were very desirable objects. The sound of the bell focused the attention on the plate which was the clue to opening the box. The auto kept the attention of the child on the problem at hand. In most cases the reward incentive seemed to be stronger than the unpleasant incentive of the sound of the bell since nearly all of the subjects overcame their dislike for the bell in order to obtain the auto. Encouragement, which was given by the experimenter whenever it seemed necessary, kept some children at the problem when they seemed about ready to give up.

COMPARISON OF THE GROUPS

The time scores for the subjects in Group A are in general much larger than those for Group B. Nine of the children in Group A and only five in Group B showed disturbance at the sound of the bell. The procedure for the two groups was identical except that no retest was given to the subjects in Group B. Although the two groups were not matched for age and I.Q., the averages for the groups show that Group B is composed of only slightly younger and of approximately equally intelligent children as those found in Group A. Furthermore, it has already been indicated that speed in accomplishing the task was not dependent upon the age or the I.Q. of the subjects but rather upon the emotional factors involved in the performance of the task. An analysis of external factors which might cause greater disturbance in Group A than in Group B was, therefore, made. Shortly before the subjects in Group A were tested in the present experiment, they were used in an experiment by Johnson. The problem involved tracing a maze with a stylus wired so that when the stylus came in contact with the side of the pathway, a bell rang and a slight electric shock was administered. The bell used was the same as the one employed in the present experiment. The subjects in Group B had had no experience with a stylus nor with the bell before the outset of the present experiment. It seems probable therefore, that there was a carry-over from the previous

experiment which acted as a confusing factor in the present experiment.

INTERPRETATION OF THE RESULTS

The speed of learning to open a simple problem box does not appear to be correlated with age or I.Q. There is a tendency for those subjects who were definitely disturbed by the sound of the bell to take longer to open the box than those subjects who were not disturbed by the bell. The difference in time scores obtained on the two groups gives indications that the apparent emotional disturbance strongly influenced the time. In general the reward incentive seemed to act more strongly than the unpleasant incentive of the sound of the bell. Encouragement tended to give the subjects self-confidence when they were discouraged with the problem. Some factors outside the control of the immediate experiment may have influenced the results, such as the success of other children in obtaining a toy.

RACIAL DIFFERENCES IN REACTION TIME AT THE PRESCHOOL LEVEL

CATHERINE HARMON[1]

During the present century there has been a great interest in racial differences in mental ability. Other aspects of racial differences have been somewhat neglected because of this consuming interest in comparative intelligence quotients. The present study was undertaken to determine whether or not racial differences are apparent in the simple reaction time of children.

Previous experiments in this field have been inconclusive. Bache (1) asserted that speed of reaction varied inversely with intelligence, though some of his observations appear not to support this view. Klineberg (3), who tested rural and urban groups of Indian, white, and negro children, maintained that racial differences on performance tests were entirely speed differences, and that these were due to environmental factors. Lambeth and Lanier (4) made an intensive study on racial differences in speed of reaction, their subjects being thirty 12-year-old whites and thirty 12-year-old negroes. Their results show no race differences in speed of simple manual movement, but as the performance becomes more complex the whites exceed the negroes in proportion to the complexity. Goodenough's (2) study of preschool children shows a low positive correlation between scores on certain intelligence tests in which speed is a factor and simple reaction time. Other researches in which sensation was the chief classification indicate that the measurement of reaction time provides a rough measure of ability in the sense category concerned.

In the present experiment, the following subjects were used:

30 Italian children, aged 3 1/2 to 6 1/2
30 Mexican children, aged 3 1/2 to 6 1/2
32 Negro children, aged 3 1/2 to 6 1/2
28 Jewish children, aged 3 1/2 to 6 1/2
13 Indian children, aged 4 1/2 to 7 1/2

Most of these children were from the lower socio-economic groups. This fact is, however, in all probability of little or no importance since Goodenough (2) found no relationship between speed of simple reaction and socio-economic status. The sexes were fairly evenly distributed at each age level.

The apparatus used was the Miles Reaction Time Board. This was placed on a low table with the child seated in front and the experimenter to his left. Preliminary trials were given until the child gave evidence of understanding what he was supposed to do. Then alternate trials were made with each hand until a total of twenty, ten with the right hand and ten with the left, had been recorded. The general procedure and instructions were the same as those used by Goodenough.

The median, expressed in sigma units, was found to be the most accurate measure of reaction speed in this test. Prolonged and premature reactions are included in the results, except in those cases in which the child's attention was obviously distracted. In these cases the trial was repeated. An auditory stimulus (sound of a buzzer) was used. At these age levels, control of the type of reaction

[1] From the Institute of Child Welfare, University of Minnesota.

CHILD DEVELOPMENT, Vol. 8, No. 3 (September, 1937)

(muscular or sensory) is practically impossible, hence the reactions probably include both types. Diffused reactions of the kind described by Luria occurred at all age levels and in all racial groups.

Results of the experiment are given in Tables 1 and 2. They show that the Italians, at all age ranges tested, react more quickly than any of the other races, and also appear to be more stable, since the average of the mean variations from the individual median is smaller. Individual differences are marked in all ages and races. The Indians were found to be slower than any of the other races. This finding is contrary to those of Bache (1) and Myers (5), and since the number of cases is so small, may easily be due to chance.

Table 1

Mean Reaction Speed* (in Sigma Units) and Mean Variations from the Mean, for Groups of Children of Six Races at Four Age Levels

	Italians			Mexicans			Negroes			Jews			Indians			Amer. Whites[1]		
Age	No.	Mean	MV	No.	Mean	MV	No.	Mean	MV	No.	Mean	MV	No.	Mean	MV	No.	Mean	MV
3 1/2	10	409	83	4	528	98	4	442	95	8	632	125	0			53	601	239
4 1/2	10	344	57	11	437	89	13	436	85	8	485	99	5	521	120	57	432	132
5 1/2	10	326	47	12	453	80	14	354	62	12	361	74	0			111	389	123
6 1/2				3	357	74	3	315	70	3	331	74	4	440	75	59	296	73

* Based upon individual means

Table 2

Mean Reaction Speed* (in Sigma Units) and Mean Variations from the Median, for Groups of Children of Six Races at Four Age Levels

	Italians			Mexicans			Negroes			Jews			Indians			Amer. Whites[1]		
Age	No.	Mean	MV	No.	Mean	MV	No.	Mean	MV	No.	Mean	MV	No.	Mean	MV	No.	Mean	MV
3 1/2	10	379	84	4	450	96	4	414	90	8	568	122	0			53	505	231
4 1/2	10	344	56	11	418	85	13	346	82	8	421	99	5	487	114	57	390	121
5 1/2	10	308	52	12	454	77	14	331	61	12	333	70	0			111	334	116
6 1/2	0			3	315	74	5	267	70	3	329	75	4	433	125	59	273	69

* Based upon individual medians

The mean variations of the Indian children were larger than those of the other races, seeming to indicate a more primitive type of reaction in the Indians. The Jewish group was slower than any except the Indians at all levels except 5 1/2 years, at which the Mexicans were slower than any group except the Indians. The Negroes rank next below the Italians in speed throughout all the age ranges. The medians indicate that the Italians were only a little speedier than the Negroes, but the Italians' mean variations from their median scores are proportionately less than the Negroes', which, as Goodenough (2) has shown, is further indication of a more mature type of reaction. The tendency to diffuse reaction, characteristic of young children, is at all ages less marked in the Italians than in the other races. The American white children studied by Goodenough consistently maintain an intermediate rank among the other ratio-racial groups included in this study. Further work on this point is desirable because of the small number of subjects in the present study.

This investigation suggests that there may be true racial differences in reac-

[1] The data on American white children are taken from Goodenough's study (2).

tion time. The Italian children showed a more mature type of reaction, age for age, than the other groups tested.

REFERENCES

(1) Bache, R. M.: Reaction time with reference to race. Psychol. Rev. 1895, 2, 474-86.

(2) Goodenough, F. L.: The development of the reactive process from early childhood to maturity. J. Exper. Psychol. 1935, 18, 431-50.

(3) Klineberg, O.: Experimental study of speed and other factors in racial differences. Arch. Psychol. 1928, 15, No. 93, p. 111.

(4) Lambeth, M., and L. Lanier: Race differences in speed of reaction. J. Genet. Psychol., 1933, 2, 255-97.

(5) Myers, C. S.: Reaction time--reports of the Cambridge anthropological expedition to Torres Straits. Vol. II, Part II, pp. 205-224.

A STUDY OF NURSES' ATTITUDES TOWARD THE BEHAVIOR PROBLEMS OF CHILDREN UNDER HOSPITAL CARE

HAZEL H. BOWLES [1]

Although Wickman, Stogdill, and others have studied the attitudes of parents and teachers toward children's problem behavior, no study has yet been made - so far as the writer has been able to ascertain - of the attitudes of nurses who care for school-age children in hospitals. The present study was carried out in an attempt to obtain data on this aspect of the question.

Previous studies of the attitudes of adults toward children's behavior problems have indicated that persons with little or no knowledge of mental hygiene tend to rank as fairly or very serious those problems that involve actions conflicting with the social code, and that they underrate the significance of withdrawing tendencies, recessive personality, and behavior traits, which mental hygienists regard as most serious. Stogdill (5) compared the ratings given various behavior problems by parents and by mental hygienists. The ratings of his mental hygienists agreed for the most part with those of the Wickman study; while parents, like teachers, were inclined to rate overt behavior problems as the more serious.

Bain (2) found that Teachers College students come nearer to the mental hygiene point of view on problem behavior after they had had one or more courses in child study. Preston and Shapler (4) comparing a group of supposedly normal grade school children with a group classed as needing psychiatric treatment found that the control group could not be distinguished from the clinic group on the basis of behavior traits. Temper tantrums, for instance, occurred in an equal percentage of both groups, while daydreaming was more prevalent in the control group than in the clinic group. This indicates the wide variance in opinion as to what different people consider as "problems". McFie (3) concluded from a study of children referred to a psychiatrist that parents tend to regard active disturbing behavior as much more serious than the personality deviations that give concern to the psychiatrist.

Only one recent investigation, that of Arlitt and Lloyd (1), has involved a study of the problem behavior exhibited by children in hospitals. These authors, however, studied preschool children, and their aim was to discover the effect of hospitalization as a factor in producing behavior problems. The object of the present study was to learn the attitudes of nurses toward various types of problem behavior, rather than to search into possible causes of the behavior. Future studies may be able to combine the two points of view and thus add to our knowledge of the cumulative and interactive effects of behavior.

Subjects and method.--It was not possible to find a group of nurses comparable in both education and term of service to answer the questions used in this study, under even partially controlled conditions. The subjects used were student nurses who had had some experience in pediatric nursing and graduate nurses actively engaged in nursing children. Some of both groups had had only high school education before their hospital training, while others had had a year or more of college work. Many of the college girls had completed a four-year course. Some

[1] From the Institute of Child Welfare, University of Minnesota.

CHILD DEVELOPMENT, Vol. 8, No. 3 (September, 1937)

nurses in both groups had had nursery school experience. Altogether there were 224 nurses from seven university hospitals. There were 83 student nurses and 17 graduates who had had some college work, and 88 students and 36 graduates who had only high school.

The questionnaire sent to the nursing schools consisted of 60 items selected from both the Wickman (6) and the Stogdill studies as likely to be applicable to children of school age in hospitals. The questionnaire is given below.

Key to Rating Scales

I. Frequency	II. Seriousness (present)	III. Seriousness (future)
a. None	a. No consequence	a. No importance at all
b. Very small number	b. Only of slight consequence	b. Only slight importance
c. In about ½ of all	c. Makes considerable difficulty	c. Considerable importance
d. In nearly all	d. Extremely grave problem	d. Extremely grave importance

1					2				3			
a	b	c	d		a	b	c	d	a	b	c	d
				1. Destroying equipment or play material								
				2. Untruthfulness (lying)								
				3. Imaginative lying								
				4. Cheating								
				5. Stealing								
				6. Profanity								
				7. Obscene notes, pictures, talk								
				8. Masturbation								
				9. Heterosexual activity (with opposite sex)								
				10. Disorderliness (violation of ward discipline)								
				11. Restlessness (overactivity) not chorea								
				12. Inattention								
				13. Lack of interest in play								
				14. Laziness								
				15. Unreliableness (irresponsible)								
				16. Disobedience								
				17. Impertinence (insubordination defiance)								
				18. Cruelty								
				19. Bullying								
				20. Quarrelsomeness								
				21. Annoying other children								
				22. Tattling								
				23. Stubbornness								
				24. Sullenness								
				25. Temper-tantrums								
				26. Impudence								
				27. Impoliteness, rudeness								
				28. Selfishness								
				29. Domineering, overbearing, dictatorial								
				30. Shyness, bashfulness								
				31. Sensitiveness								
				32. Unsociability, withdrawing								
				33. Overcritical of others								
				34. Thoughtlessness								
				35. Inquisitiveness, meddlesomeness								
				36. Silliness, "smartness," attracting attention								
				37. Unhappy, depressed, dissatisfied								
				38. Resentful								
				39. Nervousness								
				40. Fearfulness (easily frightened)								
				41. Enuresis (wetting self)								
				42. Specific fears								
				43. Dreaminess								
				44. Carelessness in personal appearance								
				45. Suspiciousness								
				46. Physical coward								
				47. Easily discouraged								
				48. Suggestible (accepts suggestion of anyone)								
				49. Disrespect for elders								
				50. Ungratefulness								
				51. Fault finding								
				52. Constant whining								
				53. Argues when corrected								
				54. Bad table manners								
				55. Demanding attention								
				56. Excessive modesty								
				57. Likes to play alone								
				58. Spends most of time reading								
				59. Asks questions continually								
				60. Unfairness								

It was not feasible to have these questionnaires filled out by all the nurses in a given hospital at the same time, since they could not all simultaneously be relieved from duty. The questionnaire was therefore given to the nurses in groups, of whom it was requested that they consult no one before making their ratings.

The data called for through the questionnaire were kept in broad categorical terms, as it was thought unlikely that the nurses would be able to make finer distinctions. In tabulating the results of the ratings, totals for each group of nurses were made for each category in each item and then combined into grand totals for all the nurses. The total of the judgments in each category was assigned the relative percentage of all judgments for that item and the category in which the median fell was determined. By making use of the Kelley-Wood Table of the Normal Probability Integral, it was possible to determine at what point on the base line of the category the median fell. In this way could be determined not only within which of the four broad categories the median for each item fell, but also what percentage of the range covered by that category fell below the median judgment. The assumption is that the true distribution of degrees of frequency or seriousness follows the form of the normal frequency curve, with the respective medians occurring at different points on a scale range from zero to the highest possible degree. The items within each category were ranked in the order of frequency and seriousness by calculating the percentage of the category at which the median number of judgments fell.

Results of the questionnaire are summarized in Tables 1, 2, 3, and 4.

Table 1

Ratings by Nurses of the Frequency and Seriousness of the Problem Behavior of Children in Hospitals

	Frequency	Present seriousness	Future seriousness
Least	Heterosexual activity Cruelty Obscene notes, pictures, and talk	Destroying equipment or play material Dreaminess Spending most of time reading Imaginative lying	Asking questions continually Restlessness Spending most of time reading Excessive modesty Inquisitiveness
Most	Demanding attention Inquisitiveness or meddlesomeness Restlessness Silliness, smartness, attracting attention Bad table manners	Stealing Temper tantrums Heterosexual activity Cruelty Unhappiness	Stealing Impudence Heterosexual activity

Table 2

Comparison of Judgments of 53 Graduate Nurses and 171 Student Nurses on: I. Frequency of Occurrence of Behavior Problems; II. Present Seriousness; III. Future Seriousness

	Graduate Nurses (N: 53)	Student Nurses (N: 171)
1. Frequency		
Obscene notes, pictures, talk	b*	a
Restlessness (overactivity)	b	a
Thoughtlessness	c	b
Silliness, "smartness," attracting attention	b	c

* a: 50% or more judge problem as occurring in none of children;
b: " " " " " " " " a very small number;
c: " " " " " " " " about 1/2 of all children.

Table 2 - Continued

Comparison of Judgments of 53 Graduate Nurses and 171 Student Nurses on: I. Frequency of Occurrence of Behavior Problems; II. Present Seriousness; III. Future Seriousness

		Graduate Nurses (N: 53)	Student Nurses (N: 171)
II.	Present Seriousness		
	Heterosexual activity	d**	c
	Unreliableness (irresponsible)	b	c
	Quarrelsomeness	b	c
	Sullenness	b	c
	Impoliteness, rudeness	b	c
	Selfishness	b	c
	Unfairness	b	c
III.	Future Seriousness		
	Destroying equipment or play material	b	c
	Imaginative lying	b	c
	Inattention	b	c
	Lack of interest in play	b	c
	Annoying other children	b	c
	Tattling	b	c
	Stubbornness	b	c
	Impudence	b	c
	Impoliteness, rudeness	b	c
	Shyness, bashfulness	b	c
	Sensitiveness	b	c
	Overcritical of others	b	c
	Thoughtlessness	b	c
	Dreaminess	b	c
	Carelessness in personal appearance	b	c
	Ungratefulness	b	c
	Bad table manners	b	c
	Demanding attention	b	c

** b, rating indicates only slight importance; c, considerable importance; d, extremely grave importance.

Table 3

Comparison of Judgments of Nurses with High School Education and Judgments of Nurses with One Year or More of College Education on: I. Frequency; II. Present Seriousness; III. Future Seriousness

		High School (N: 124)	College (N: 100)
I.	Frequency		
	Obscene notes, pictures, talk	b	a
	Restlessness (overactivity) not chorea	b	c
	Thoughtlessness	b	c
	Silliness, "smartness," attracting attention	b	c
	Bad table manners	b	c
	Demanding attention	b	c
II.	Present Seriousness		
	Profanity	c	b
	Quarrelsomeness	c	b
	Sullenness	c	b
	Impoliteness, rudeness	c	b
	Selfishness	c	b
	Resentfulness	c	b
	Easily discouraged	c	b
	Disrespect for elders	c	b
	Argues when corrected	c	b
	Bad table manners	c	b
III.	Future Seriousness		
	Untruthfulness (lying)	c	d
	Cheating	c	d
	Annoying other children	b	c
	Tattling	b	c
	Impudence	c	b
	Shyness, bashfulness	c	b
	Dreaminess	b	c
	Demanding attention	b	c

Table 4

Comparison of Judgments of Nurses with and without Nursery School Experience on: I. Frequency; II. Present Seriousness; III. Future Seriousness

		Nursery School (N: 63)	No Nursery School (N: 161)
I.	Frequency		
	Stealing	a	b
	Obscene notes, pictures, talk	a	b
	Domineering, overbearing, dictatorial	c	b
	Sensitiveness	c	b
	Thoughtlessness	c	b
	Silliness, "smartness," attracting attention	c	b
	Easily discouraged	c	b
	Bad table manners	c	b
II.	Present Seriousness		
	Profanity	b	c
	Masturbation	b	c
	Impertinence (insubordination, defiance)	b	c
	Quarrelsomeness	b	c
	Annoying other children	b	c
	Stubbornness	b	c
	Sullenness	b	c
	Impoliteness, rudeness	b	c
	Selfishness	b	c
	Unsociability, withdrawing	b	c
	Resentfulness	b	c
	Nervousness	b	c
	Fearfulness (easily frightened)	b	c
	Physical coward	b	c
	Easily discouraged	b	c
	Disrespect for elders	b	c
	Unfairness	b	c
III.	Future Seriousness		
	Heterosexual activity	c	d
	Inattention	c	b
	Annoying other children	c	b
	Tattling	c	b
	Temper tantrums	b	c
	Impudence	b	c
	Impoliteness, rudeness	b	c
	Shyness, bashfulness	b	c
	Dreaminess	c	b
	Bad table manners	b	c

On all items not specifically referred to in the tables, the two groups were in agreement.

The following comparisons were also made, with results as indicated:

A. Judgments of present seriousness of problems by teachers[1] and by nurses:

Nurses rated as of more present seriousness than teachers did, tattling, unsociability, fearfulness, stubbornness; depression, enuresis, and temper tantrums. Teachers rated as more serious carelessness in personal appearance, laziness, impudence, impoliteness, lack of interest, destroying equipment, impertinence, and obscenity.

B. Judgments of present seriousness of problems by parents[2] and by nurses.

Nurses rated as more serious domineering, stubborness, being easily discouraged, constant whining, specific fears, depression, and enuresis. Parents rated as more serious than nurses problems associated with ungratefulness, suggestibility, physical cowardice, destructiveness, obscenity, unreliability, and masturbation.

The correlation between judgments of teachers and nurses is .68 ± .05; that between judgments of parents and nurses is .77 ± .03.

[1] In Wickman study.

[2] In Stogdill study.

C. Judgments of future seriousness of problems by mental hygienists and by nurses.

1. On the Wickman study, mental hygienists judged as more serious silliness, tattling, lack of interest in play, stubbornness, dreaminess, shyness, sullenness, sensitiveness, overcriticism, bullying, fearfulness, resentfulness, suspiciousness, and unsociability. Nurses of the present study rated as more serious disorderliness, disobedience, masturbation, impertinence, laziness, carelessness in personal appearance, obscenity, enuresis, heterosexual activity, cheating, untruthfulness, unreliability, temper tantrums, and stealing. The correlation is .34 ± .09.

2. On the Stogdill study, mental hygienists rated as more serious than nurses did problems associated with restlessness, tattling, demanding attention, liking to play alone, fault-finding, lack of interest in play, destructiveness, excessive modesty, nervousness, suspiciousness, constant whining, and specific fears. Nurses rated as more serious than mental hygienists did, bad table manners, carelessness in personal appearance, disrespect for elders, overcriticism, argumentativeness, impoliteness, laziness, domineering, masturbation, disobedience, and heterosexual activity. The correlation in this case was .56 ± .06.

D. Judgment of present seriousness of problems by mental hygienists and nurses.

1. On the Wickman study, mental hygienists rated as more serious suspiciousness, shyness, dreaminess, sensitiveness, suggestibility, unsociability, resentfulness, and overcriticism. Nurses rated as more serious than mental hygienists did, disobedience, disorderliness, masturbation, impertinence, heterosexual activity, enuresis, cheating, temper tantrums, lying, and stealing. The correlation is .40 ± .08.

1. On the Stogdill study, mental hygienists rated as more serious destructiveness, excessive modesty, suspiciousness, liking to play alone, lack of interest in play, unsociability, dreaminess, physical cowardice, resentfulness, irresponsibility, and specific fears. Nurses rated as more serious disrespectfulness, argumentativeness, impertinence, masturbation, disobedience, heterosexual activity, bad table manners, stubbornness, domineering, and impoliteness. The correlation here is .59 ± .06.

It was reasonable to suppose that nurses would be more likely to rank as least frequent those problem tendencies that came least often to their attention, and as least serious those that caused the smallest amount of disturbance in the ward. This assumption was borne out by the results. There are observable, however, certain notable divergences from this situation.

The main conclusions of this study may be summed up as follows:

1. Nurses rate as more frequent problems associated with an effort to get attention, and affecting only the welfare of the individual child. Transgressions against commonly accepted moral regulations are least frequently reported.

2. Problems that conflict with the conventional code, and introvert tendencies, are considered by nurses as most serious both at the time they occur and for the future welfare of the child.

3. Disciplinary offences are judged by nurses to be of considerable importance to the child's future welfare.

4. Attitudes of nurses toward behavior problems of children resemble most closely those of the parents studied by Stogdill.

5. Attitudes of nurses toward the future seriousness of children's behavior

problems resemble more closely those of the mental hygienists in the Stogdill study than those of the mental hygienists in the Wickman study.

6. Attitudes of nurses are more similar to attitudes of parents in the Stogdill study than to those of mental hygienists in the same study.

7. Few of the problems that parents and teachers meet are noted in as many as one-half the children of school age who are under hospital care.

8. Student nurses consider behavior that disturbs the peace of the ward, withdrawing tendencies, and demanding attention, as more serious than graduates do. Graduates rate heterosexual activity as very serious at the time of occurrence.

9. Nurses who have had only high school education believe the most frequent problems to be those in which the child is seeking for attention. The college group reports these problems, also those that interfere with their authority or with the happiness of others, as more serious for the present than the high school group does. The high school group sees greater seriousness for the future in problems that cause conflict with the integrity of the group.

10. There is a tendency for nurses without nursery school experience to rate more problems as serious for the present welfare of the child than those with nursery school experience do. These problems include withdrawing traits, conflicts with the conventional codes, and conflicts with authority. When considering the future welfare of the child, the nurses without nursery school experience consider opposition to authority, heterosexual activity, and shyness more serious.

11. If the attitudes of the mental hygienists in these studies are correct, training schools for nurses need to place more emphasis on mental hygiene in the curriculum.

REFERENCES

(1) Arlitt, A. H., and Lloyd, W.: Hospitalization as a factor in producing behavior problems. Ninth International Congress of Psych. Proc. and Papers, 62-63.

(2) Bain, W. E.: A study of the attitudes of teachers toward behavior problems. Child Development, 1934, 5, 19-35.

(3) McFie, B. S.: Behavior and personality difficulties in school children. Brit. J. Educ. Psychol., 1934, 4, part 1, 30-46.

(4) Preston, G. H., and Shapler, W. McL.: A study of problems of 'normal' children. Am. J. Orthopsychiatry, 1931, 1, 245-256.

(5) Stogdill, R. M.: Parental attitudes and mental hygiene standards. Ment. Hyg., 1931, 15, 813-827.

(6) Wickman, E. K.: Children's Behavior and Teachers' Attitudes. New York, Commonwealth Fund, 1928, pp. 1-247.